tsvwg
Internet-Draft
Intended status: Standards Track
Expires: May 5, 2021
Z. Zhang
R. Bonica
K. Kompella
Juniper Networks
November 01, 2020

Generic Transport Functions
draft-zzhang-tsvwg-generic-transport-functions-00

Abstract

Some functionalities (e.g. fragmentation/reassembly and Encapsulating Security Payload) provided by IPv6 can be viewed as independent of IPv6 or even IP entirely. This document proposes to provide those functionalities at different layers (e.g., MPLS, BIER or even Ethernet) independent of IP.

Table of Contents

1. Introduction

Consider an operator providing Ethernet services such as pseudowires, VPLS or EVPN. The Ethernet frames that a Provider Edge (PE) device receives from a Customer Edge (CE) device may have a larger size than the PE-PE path MTU (pMTU) in the provider network. This could be because

1. the provider network is built upon virtual connections (e.g. pseudowires) provided by another infrastructure provider, or

2. the customer network uses jumbo frames while the provider network does not, or

3. the provider-side overhead for transporting customers packets across the network pushes past the pMTU.

In any case, the provider simply cannot require its customers to change their MTU.

To get those large frames across the provider network, currently the only workaround is to encapsulate the frames in IP (with or without GRE) and then fragment the IP packets. Even if MPLS is used for

service delimiting, IP is used for transporation (MPLS over IP/GRE). This may not be desirable in certain deployment scenarios, where MPLS is the preferred transport or IP encapsulation overhead is deemed excessive.

IPv6 fragmentation and reassembly are based on the IPv6 Fragmentation header below [RFC8200]:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Next Header  |   Reserved    |      Fragment Offset    |Res|M|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Identification                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: IPv6 Fragmentation Header

This document proposes reusing this header in non-IP contexts, since the fragmentation/reassembly function is actually independent of IPv6 except the following aspects:

- o The fragment header is identified as such by the "previous" header.
- o The "Next Header" value is from the "Internet Protocol Numbers" registry.
- o The "Identification" value is unique in the (source, destination) context provided by the IPv6 header

The "Identification" field, in conjunction with the IPv6 source and destination identifies fragments of the original packet, for the purpose of reassembly.

Therefore, the fragmentation/reassembly function can be applied at other layers as long as a) the fragment header is identified as such; and b) the context for packet identification is provided. Examples of such layers include MPLS, BIER, and Ethernet (if IEEE determines it is so desired).

For the layers where the IETF is concerned, the "Next Header" value will still be from the "Internet Protocol Numbers" registry when the

function is applied at non-IP layers.

For the same consideration, the IP Encapsulating Security Payload (ESP) [RFC4303] could also be applied at other layers if ESP is desired there. For example, if for whatever reason the Ethernet service provider wants to provide ESP between its PEs, it could do so without requiring IP encapsulation if ESP is applied at non-IP layers.

The possibility of applying some other IP functions (e.g. Authentication Header [RFC4302]) is for further study.

2. Specifications

2.1. Generic Fragmentation Header

For generic fragmentation/reassembly functionality independent of IP, the following Generic Fragmentation Header (GFH) is defined:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Next Header  | Header Length |      Fragment Offset    |R|S|M|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Identification                        |
|                           (variable)                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: Generic Fragmentation Header

The "Next Header", "Fragment Offset" and "M" flag bit fields are as in the IPv6 Fragmentation Header.

Header Length: the number of octets of the entire header.

R: The "R" flag bit is reserved. It MUST be 0 on transmitting and ignored on receiving.

Identification: at least 4-octet long.

S: If the "S" flag bit is clear, the context for the Identification field is provided by the outer header, and only the source-identifying information in the outer header is used. If the "S"

flag bit is set, the variable Identification field encodes both source-identifying information (e.g. the IP address of the node adding the GFH) and an identification number unique within that source.

The outer header MUST identify that a Generic Fragmentation Header follows and MAY carry source-identifying information.

If the outer header is BIER, a TBD value for the "proto" field in the BIER header identifies that a GFH follows. If the "S" flag bit is clear, the "BFIR-id" field in the BIER header provides the context for the "Identification" field.

If the outer header is MPLS, the "S" flag bit MAY be clear if the the label preceeding the GFH identifies the sending BFR in addition to indicating that a GFH follows (see Section 2.2).

2.2. MPLS Signaling

When GFH is used with MPLS, the preceeding label needs to indicate that a GFH follows, and optionally identify the node that does the fragmentation. The label can be signaled via BGP or IGP as sepcified below.

2.2.1. BGP Signaling

This document defines a new transitive BGP "GFH Labels" attribute, very similar to the "PE Distinguisher Labels" attribute defined in [RFC6514] (and the text below is adapted from Section 8 of [RFC6514]):

```
+---------------------------------+
|     Node Address                |
+---------------------------------+
|     Label (3 octets)            |
+---------------------------------+
.......
+---------------------------------+
|     Node Address                |
```

```
+--------------------------------+
|     Label (3 octets)           |
+--------------------------------+
```

The Label field contains an MPLS label encoded as 3 octets, where the high-order 20 bits contain the label value. The Node Address MAY be 0, meaning that the following label only indicates a GFH follows when the label is used in the label stack of a data packet.

The Node Address MAY also be a unicast address, indicating that the following label when used in the label stack of a data packet will both indicate that a GFH follows and identify the sending node.

If a node supports GFH with MPLS, it attaches the attribute in the BGP routes for its local addresses. A border router SHOULD remove the attribute if no node beyond the border will use GFH with MPLS to send traffic to the corresponding addresses.

A router that supports the attribute considers this attribute to be malformed if the Node Address field does not contain a unicast address or 0. The attribute is also considered to be malformed if: (a) the Node Address field is expected to be an IPv4 address, and the length of the attribute is not a multiple of 7 or (b) the Node Address field is expected to be an IPv6 address, and the length of the attribute is not a multiple of 19. The Address Family Indicator (AFI) of the BGP route that the attribute is attached to provides the

information on whether the Node Address field contains an IPv4 or IPv6 address. Each of the Node Addresses in the attribute MUST be of the same address family as the route that is carrying the attribute.

2.2.2. IGP Signaling

This document defines an OSPFv2 "GFH Labels" sub-TLV of OSPFv2 Extended Prefix TLV [RFC7684], with the value part being the same as BGP "GFH Labels" attribute above. If an OSPFv2 router surports GFH with MPLS, it includes the GFH Labels sub-TLV in the Extended Prefix TLV that is attached to its local addresses advertised in its OSPFv2 Extended Prefix Opaque LSA.

Similary, This document defines an OSPFv3 "GFH Labels" sub-TLV of OSPFv3 Intra/Inter-Area-Prefix TLVs [RFC8362], with the value part

being the same as BGP "GFH Labels" attribute above. If an OSPFv3 router surports GFH with MPLS, it includes the GFH Labels sub-TLV in the Intra-Area-Prefix TLV for its local addresses.

This document also defines an ISIS "GFH Labels" sub-TLV of ISIS prefix-reachability TLV [RFC5120] [RFC5305] [RFC5308], with the value part being the same as BGP "GFH Labels" attribute above. If an ISIS router surports GFH with MPLS, it includes the sub-TLV to the prefix-reachability TLV for its local addresses.

For both OSPF and ISIS, when advertising a prefix from one area/level to another, if there is a "GFH Labels TLV" attached in the source area/level, the TLV SHOULD be attached in the target area/level and the prefix SHOULD NOT be summarized.

2.3. Generic ESP/Authentication Header

To be specified in future revisions.

3. Security Considerations

To be provided.

4. IANA Considerations

This document makes the following IANA requests:

- o A new BGP Attribute type for "GFH Labels" from the BGP Path Attributes registry

- o A new OSPFv2 sub-TLV type for "GFH Labels" from the OSPFv2 Extended Prefix TLV Sub-TLVs registry

- o A new OSPFv3 sub-TLV type for "GFH Labels" from the OSPFv3 Extended-LSA sub-TLV registry

- o A new BIER Next Protocol Identifier value for GFH from BIER Next Protocol Identifiers registry

5. Acknowledgements

6. References

6.1. Normative References

[RFC4303] Kent, S., "IP Encapsulating Security Payload (ESP)", RFC 4303, DOI 10.17487/RFC4303, December 2005, <https://www.rfc-editor.org/info/rfc4303>.

[RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, DOI 10.17487/RFC5120, February 2008, <https://www.rfc-editor.org/info/rfc5120>.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <https://www.rfc-editor.org/info/rfc5305>.

[RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, DOI 10.17487/RFC5308, October 2008, <https://www.rfc-editor.org/info/rfc5308>.

[RFC7684] Psenak, P., Gredler, H., Shakir, R., Henderickx, W., Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute Advertisement", RFC 7684, DOI 10.17487/RFC7684, November 2015, <https://www.rfc-editor.org/info/rfc7684>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[RFC8362] Lindem, A., Roy, A., Goethals, D., Reddy Vallem, V., and F. Baker, "OSPFv3 Link State Advertisement (LSA) Extensibility", RFC 8362, DOI 10.17487/RFC8362, April 2018, <https://www.rfc-editor.org/info/rfc8362>.

6.2. Informative References

[RFC4302] Kent, S., "IP Authentication Header", RFC 4302, DOI 10.17487/RFC4302, December 2005, <https://www.rfc-editor.org/info/rfc4302>.

[RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, DOI 10.17487/RFC6514, February 2012, <https://www.rfc-editor.org/info/rfc6514>.

Authors' Addresses

Zhaohui Zhang
Juniper Networks
1133 Innovation Way
Sunnyvale 94089
USA

Phone: +1 408 745 2000
Email: zzhang@juniper.net

Ron Bonica
Juniper Networks
1133 Innovation Way
Sunnyvale 94089
USA

Phone: +1 408 745 2000
Email: rbonica@juniper.net

Kireeti Kompella
Juniper Networks
1133 Innovation Way
Sunnyvale 94089
USA

Phone: +1 408 745 2000
Email: kireeti@juniper.net