Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 24, 2014

E. Crabbe, Ed.
Google
L. Yong, Ed.
Huawei USA
X. Xu, Ed.
Huawei Technologies

October 21, 2013

Generic UDP Encapsulation for IP Tunneling
draft-yong-tsvwg-gre-in-udp-encap-02

Abstract

This document describes a method of encapsulating arbitrary protocols within GRE and UDP headers. In this encapsulation, the source UDP port may be used as an entropy field for purposes of loadbalancing while the payload protocol may be identified by the GRE Protocol Type.

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].

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Load balancing, or more specifically, statistical multiplexing of traffic using Equal Cost Multi-Path (ECMP) and/or Link Aggregation Groups (LAGs) in IP networks is a widely used technique for creating higher capacity networks out of lower capacity links. Most existing routers in IP networks are already capable of distributing IP traffic flows over ECMP paths and/or LAGs on the basis of a hash function performed on flow invariant fields in IP packet headers and their payload protocol headers. Specifically, when the IP payload is a User Datagram Protocol (UDP)[RFC0768] or Transmission Control Protocol (TCP) packet, router hash functions frequently operate on the five-tuple of the source IP address, the destination IP address, the source port, the destination port, and the protocol/next-header

Several tunneling techniques are in common use in IP networks, such as Generic Routing Encapsulation (GRE) [RFC2784], MPLS [RFC4023] and L2TPv3 [RFC3931]. GRE is an increasingly popular encapsulation choice, especially in environments where MPLS is unavailable or unnecessary. Unfortunately, use of common GRE endpoints may reduce the entropy available for use in load balancing, especially in environments where the GRE Key field [RFC2890] is not readilly available for use as entropy in forwarding decisions.

This document defines a generic GRE-in-UDP encapsulation for tunneling arbitrary network protocol payloads across an IP network environment where ECMP or LAGs are used. The GRE header provides payload protocol de-multiplexing by way of it’s protocol type field [RFC2784] while the UDP header provides additional entropy by way of it’s source port.

This encapsulation method requires no changes to the transit IP network. Hash functions in most existing IP routers may utilize and benefit from the use of a GRE-in-UDP tunnel without needing any change or upgrade to to their ECMP implementations. The encapsulation mechanism is applicable to a variety of IP networks including Data Center and wide area networks.

2. Terminology

The terms defined in [RFC0768] are used in this document.

3. Procedures

When a tunnel ingress device conforming to this document receives a packet, the ingress MUST encapsulate the packet in UDP and GRE headers and set the destination port of the UDP header to [TBD] Section 6. he ingress device must also insert the payload protocol type in the GRE Protocol Type field. The ingress device SHOULD set the UDP source port based on flow invariant fields from the payload header, otherwise it should be set to a randomly selected constant value, e.g. zero, to avoid packet flow reordering. How a tunnel ingress generates entropy from the payload is outside the scope of this document. The tunnel ingress MUST encode its own IP address as the source IP address and the egress tunnel endpoint IP address. The TTL field in the IP header must be set to a value appropriate for delivery of the encapsulated packet to the tunnel egress endpoint.

When the tunnel egress receives a packet, it must remove the outer UDP and GRE headers. Section 5 describes the error handling when this entity is not instantiated at the tunnel egress.

To simplify packet processing at the tunnel egress, packets destined to this assigned UDP destination port [TBD] SHOULD have their UDP checksum and Sequence flags set to zero because the egress tunnel only needs to identify this protocol. Although IPv6 [RFC2460] restricts the processing a packet with the UDP checksum of zero, [RFC6935] and [RFC6936] relax this constraint to allow the zero UDP checksum.

The tunnel ingress may set the GRE Key Present, Sequence Number Present, and Checksum Present bits and asscociated fields in the GRE header defined by [RFC2784] and [RFC2890].

In addition IPv6 nodes MUST conform to the following:

1. the IPv6 tunnel ingress and egress SHOULD follow the node requirements specified in Section 4 of [RFC6936] and the usage requirements specified in Section 5 of [RFC6936]

2. IPv6 transit nodes SHOULD follow the requirements 9, 10, 11 specified in Section 5 of [RFC6936].

The format of the GRE-in-UDP encapsulation for both IPv4 and IPv6 outer headersis shown in the followingfigures:

```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

IPv4 Header:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version|  IHL  |Type of Service|          Total Length         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Identification        |Flags|      Fragment Offset    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Time to Live |Protcol=17[UDP]|          Header Checksum      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Source IPv4 Address                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Destination IPv4 Address                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

UDP Header:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Source Port = XXXX      |       Dest Port = TBD         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           UDP Length          |        UDP Checksum           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

GRE Header:
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|C| |K|S| Reserved0       | Ver |         Protocol Type         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Checksum (optional)      |       Reserved1 (Optional)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Key (optional)                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Sequence Number (Optional)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: UDP+GRE IPv4 headers

```
   0                   1                   2                   3
   0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

   IPv6 Header:
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Version| Traffic Class |           Flow Label                  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |         Payload Length        | NxtHdr=17[UDP]|   Hop Limit   |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   +                                                               +
   |                                                               |
   +                     Outer Source IPv6 Address                 +
   |                                                               |
   +                                                               +
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   +                                                               +
   |                                                               |
   +                  Outer Destination IPv6 Address               +
   |                                                               |
   +                                                               +
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+


   UDP Header:
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |       Source Port = XXXX      |       Dest Port = TBD         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           UDP Length          |        UDP Checksum           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   GRE Header:
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |C| |K|S| Reserved0       | Ver |         Protocol Type         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |      Checksum (optional)      |       Reserved1 (Optional)    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                         Key (optional)                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                 Sequence Number (Optional)                    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: UDP+GRE IPv6 headers

The total overhead increase for a UDP+GRE tunnel without use of optional GRE fields, representing the lowest total overhead increase, is 32 bytes in the case of IPv4 and 52 bytes in the case of IPv6. The total overhead increase for a UDP+GRE tunnel with use of GRE Key, Sequence and Checksum Fields, representing the highest total overhead increase, is 44 bytes in the case of IPv4 and 64 bytes in the case of IPv6.

4. Encapsulation Considerations

GRE-in-UDP encapsulation allows the tunneled traffic to be unicast, broadcast, or multicast traffic. Entropy may be generated from the header of tunneled unicast or broadcast/multicast packets at tunnel ingress. The mapping mechanism between the tunneled multicast traffic and the multicast capability in the IP network is transparent and independent to the encapsulation and is outside the scope of this document.

If tunnel ingress must perform fragmentation on a packet before encapsulation, it MUST use the same source UDP port for all packet fragments. This ensures that the transit routers will forward the packet fragments on the same path. GRE-in-UDP encapsulation introduces some overhead as mentioned in section 3, which reduces the effective Maximum Transmission Unit (MTU) size. An operator should factor in this addition overhead bytes when considering an MTU size for the payload to reduce the likelihood of fragmentation.

To ensure the tunneled traffic gets the same treatment over the IP network, prior to the encapsulation process, tunnel ingress should process the payload to get the proper parameters to fill into the IP header such as DiffServ [[RFC2983]]. Tunnel end points that support ECN MUST use the method described in [RFC6040] for ECN marking propagation. This process is outside of the scope of this document.

Note that the IPv6 header [RFC2460] contains a flow label field that may be used for load balancing in an IPv6 network [RFC6438]. Thus in an IPv6 network, either GRE-in-UDP or flow labels may be used in order to improve load balancing performance. Use of GRE-in-UDP encapsulation provides a unified hardware implementation for load balancing in an IP network independent of the IP version(s) in use.

5. Backward Compatibility

It is assumed that tunnel ingress routers must be upgraded in order to support the encapsulations described in this document.

No change is required at transit routers to support forwarding of the encapsulation described in this document.

If a router that is intended for use as a tunnel egress does not support the GRE-in-UDP encapsulation described in this document, it will not be listening on destination port [TBD]. In these cases, the router will conform to normal UDP processing and respond to the tunnel ingress with an ICMP message indicating "port unreachable" according to [RFC0792]. Upon receiving this ICMP message, the tunnel ingress MUST NOT continue to use GRE-in-UDP encapsulation toward this tunnel egress without management intervention.

6. IANA Considerations

IANA is requested to make the following allocation: Service Name: GRE-in-UDP Transport Protocol(s): UDP Assignee: IESG iesg@ietf.org Contact: IETF Chair chair@ietf.org Description: GRE-in-UDP Encapsulation Reference: [This.I-D] Port Number: TBD Service Code: N/A Known Unauthorized Uses: N/A Assignment Notes: N/A

7. Security Considerations

7.1. Vulnerability

Neither UDP nor GRE encapsulation effects security for the payload protocol. When using GRE-in-UDP, Network Security in a network is similar to that of a network using GRE.

Use of ICMP for signaling of the GRE-in-UDP encapsulation capability adds a security concern. Tunnel ingress devices may want to validate the origin of ICMP Port Unreachable messages before taking action. The mechanism for performing this validation is out of the scope of this document.

In an instance where the UDP src port is not set based et the flow invariant fields from the payload header, a random port SHOULD be selected in order to minimize the vulnerability to off-path attacks. [RFC6056] How the src port randomization occurs is outside scope of this document.

8. Acknowledgements

The Authors would like to thank Vivek Kumar, Ron Bonica, Joe Touch, Ruediger Geib, Gorry Fairhurst, and David Black for their review and valuable input on this draft.

9. Contributing Authors

The following people all contributed significantly to this document and are listed below in alphabetical order:

John E. Drake
Juniper Networks

Email: jdrake@juniper.net

Adrian Farrel
Juniper Networks

Email: adrian@olddog.co.uk

Vishwas Manral
Hewlett-Packard Corp.
3000 Hanover St, Palo Alto.

Email: vishwas.manral@hp.com

Carlos Pignataro
Cisco Systems
7200-12 Kit Creek Road
Research Triangle Park, NC 27709 USA

EMail: cpignata@cisco.com

Yongbing Fan
China Telecom
Guangzhou, China.
Phone: +86 20 38639121

10. References

10.1. Normative References

[RFC0768] Postel, J., "User Datagram Protocol", STD 6, RFC 768, August 1980.

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.

[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.

[RFC2983] Black, D., "Differentiated Services and Tunnels", RFC 2983, October 2000.

[RFC5405] Eggert, L. and G. Fairhurst, "Unicast UDP Usage Guidelines for Application Designers", BCP 145, RFC 5405, November 2008.

[RFC6040] Briscoe, B., "Tunnelling of Explicit Congestion Notification", RFC 6040, November 2010.

[RFC6335] Cotton, M., Eggert, L., Touch, J., Westerlund, M., and S. Cheshire, "Internet Assigned Numbers Authority (IANA) Procedures for the Management of the Service Name and Transport Protocol Port Number Registry", BCP 165, RFC 6335, August 2011.

[RFC6438] Carpenter, B. and S. Amante, "Using the IPv6 Flow Label for Equal Cost Multipath Routing and Link Aggregation in Tunnels", RFC 6438, November 2011.

[RFC6935] Eubanks, M., Chimento, P., and M. Westerlund, "IPv6 and UDP Checksums for Tunneled Packets", RFC 6935, April 2013.

[RFC6936] Fairhurst, G. and M. Westerlund, "Applicability Statement for the Use of IPv6 UDP Datagrams with Zero Checksums", RFC 6936, April 2013.

10.2. Informative References

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, September 1981.

[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.

[RFC4023] Worster, T., Rekhter, Y., and E. Rosen, "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", RFC 4023, March 2005.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

[RFC4884] Bonica, R., Gan, D., Tappan, D., and C. Pignataro, "Extended ICMP to Support Multi-Part Messages", RFC 4884, April 2007.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, November 2012.

Authors' Addresses

Edward Crabbe (editor)
Google
1600 Amphitheatre Parkway
Mountain View, CA 94102
US

Email: edward.crabbe@gmail.com

Lucy Yong (editor)
Huawei USA
5340 Legacy Drive
San Jose, TX 75025
US

Email: lucy.yong@huawei.com

Xiaohu Xu (editor)
Huawei Technologies
Beijing
China

Email: xuxiaohu@huawei.com