

INTAREA Working Group
Internet-Draft
Intended status: Standards Track
Expires: June 24, 2021

X. Xu
Alibaba Inc
H. Assarpour
Broadcom
S. Ma
Mellanox
D. Bernier
Canada Bell
D. Dukes
Cisco
S. Hegde
Juniper
Y. Lee
Comcast
Y. Fan
China Telecom
December 21, 2020

Encapsulating IP in UDP
draft-xu-intarea-ip-in-udp-10

Abstract

Existing IP-in-IP encapsulation technologies are not adequate for efficient load balancing of IP-in-IP traffic across IP networks. This document specifies additional IP-in-IP encapsulation technology, referred to as IP-in-UDP (User Datagram Protocol), which can facilitate the load balancing of IP-in-IP traffic across IP networks.

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](#)].

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 <https://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 June 24, 2021.

Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	2
2.	Terminology	4
3.	Encapsulation in UDP	4
4.	Processing Procedures	5
5.	Congestion Considerations	6
6.	Applicability Statements	6
7.	Acknowledgements	7
8.	IANA Considerations	7
9.	Security Considerations	7
10.	References	8
10.1.	Normative References	8
10.2.	Informative References	9
	Authors' Addresses	10

[1.](#) Introduction

To fully utilize the bandwidth available in IP networks and/or facilitate recovery from a link or node failure, load balancing of traffic over Equal Cost Multi-Path (ECMP) and/or Link Aggregation Group (LAG) across IP networks including Wide Area Networks (WAN) and Data Center Networks (DCN) is widely used. There are increasing applications of the IP-in-IP encapsulation on the WAN and DCN environments, such as Global Accelerator (GA) on public cloud providers' WAN and high-performance DCNs. [\[RFC5640\]](#) describes a method for improving the load balancing efficiency of IP-in-IP traffic over Layer Two Tunneling Protocol - Version 3 (L2TPv3)

[RFC3931] and Generic Routing Encapsulation (GRE) [RFC2784] encapsulations. However, this method requires core routers or switches to perform hash calculation on the "load-balancing" field contained in tunnel encapsulation headers (i.e., the Session ID field in L2TPv3 headers or the Key field in GRE headers), which is not widely supported by existing core routers and data center switches.

Most existing routers and data center switches are already capable of distributing IP traffic "microflows" [RFC2474] over ECMP paths and/or LAG based on the hash of the five-tuple of User Datagram Protocol (UDP) [RFC0768] and Transmission Control Protocol (TCP) packets (i.e., source IP address, destination IP address, source port, destination port, and protocol). By encapsulating the IP traffic into an UDP tunnel and using the source port of the UDP header as an entropy field, the existing load-balancing capability as mentioned above can be leveraged to provide fine-grained load-balancing of IP-in-IP traffic over IP networks. This is similar to why LISP [RFC6830], MPLS-in-UDP [RFC7510] and VXLAN [RFC7348] use UDP encapsulation. Therefore, this specification defines an IP-in-UDP encapsulation method which is an alternative encapsulation used in [RFC5565] in addition to L2TPv3 and GRE.

IPv6 flow label has been proposed as an entropy field for load balancing in IPv6 network environment [RFC6438]. However, as stated in [RFC6936], the end-to-end use of flow labels for load balancing is a long-term solution and therefore the use of load balancing using the transport header fields would continue until any widespread deployment is finally achieved. As such, IP-in-UDP encapsulation would still have a practical application value in the IPv6 networks during this transition timeframe. Of course, it RECOMMENDED that the IPv6 flow label is filled with an entropy value as well. In this way, core routers or switches could perform load-balancing of IP-in-IP traffic using either the approach as described in [RFC6438] or the UDP five tuple accordingly.

Similarly, the IP-in-UDP encapsulation format defined in this document by itself cannot ensure the integrity and privacy of data packets being transported through the IP-in-UDP tunnels and cannot enable the tunnel decapsulators to authenticate the tunnel encapsulator. Therefore, in the case where any of the above security issues is concerned, the IP-in-UDP SHOULD be secured with DTLS [RFC6347]. For more details, please see [Section 9](#) of Security Considerations.

2. Terminology

3. Encapsulation in UDP

IP-in-UDP encapsulation format is shown as follows:

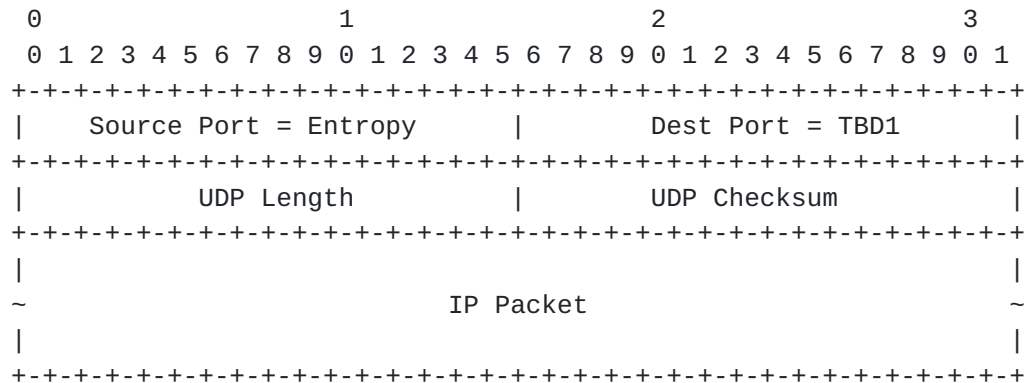


Figure 1: IP-in-UDP Encapsulation Format

Source Port of UDP:

This field contains a 16-bit entropy value that is generated by the encapsulator to uniquely identify a flow. What constitutes a flow is locally determined by the encapsulator and therefore is outside the scope of this document. What algorithm is actually used by the encapsulator to generate an entropy value is outside the scope of this document.

To ensure that the source port number is always in the range 49152 to 65535 (Note that those ports less than 49152 are reserved by IANA to identify specific applications/protocols) which may be required in some cases, instead of calculating a 16-bit entropy, the encapsulator SHOULD calculate a 14-bit entropy and use those 14 bits as the least significant bits of the source port field while the most significant two bits SHOULD be set to binary 11. That still conveys 14 bits of entropy information which would be enough as well in practice.

Destination Port of UDP:

This field is set to a value (TBD1) allocated by IANA to indicate that the UDP tunnel payload is an IP packet. As for whether the encapsulated IP packet is IPv4 or IPv6, it would be

determined according to the Version field in the IP header of the encapsulated IP packet.

UDP Length:

The usage of this field is in accordance with the current UDP specification [[RFC0768](#)].

UDP Checksum:

For IPv4 UDP encapsulation, this field is RECOMMENDED to be set to zero for performance or implementation reasons because the IPv4 header includes a checksum and use of the UDP checksum is optional with IPv4. For IPv6 UDP encapsulation, the IPv6 header does not include a checksum, so this field MUST contain a UDP checksum that MUST be used as specified in [[RFC0768](#)] and [[RFC2460](#)] unless one of the exceptions that allows use of UDP zero-checksum mode (as specified in [[RFC6935](#)]) applies.

IP Packet:

This field contains one IP packet.

4. Processing Procedures

This IP-in-UDP encapsulation causes the original packets to be forwarded across a transit IP network via "UDP tunnels". While performing IP-in-UDP encapsulation, tunnel encapsulators would generate an entropy value and encode it in the Source Port field of the UDP header. The Destination Port field is set to a value (TBD1) allocated by IANA to indicate that the UDP tunnel payload is an IP packet. Transit routers or switches, upon receiving these UDP encapsulated packets, could balance these packets based on the hash of the five-tuple of UDP packets. Tunnel decapsulators receiving these UDP encapsulated packets MUST decapsulate these packets by removing the UDP header and then forward them accordingly (assuming that the Destination Port was set to the reserved value pertaining to IP).

Similar to all other IP-in-IP tunneling technologies, IP-in-UDP encapsulation introduces overheads and reduces the effective Maximum Transmission Unit (MTU) size. IP-in-UDP encapsulation may also

impact Time-to-Live (TTL) or Hop Count (HC) and Differentiated Services (DSCP). Hence, IP-in-UDP MUST follow the corresponding procedures defined in [[RFC2003](#)].

Tunnel encapsulators MUST NOT fragment UDP encapsulated IP packets, and when the outer IP header is IPv4, tunnel encapsulators MUST set the DF bit in the outer IPv4 header. It is strongly RECOMMENDED that the transit IP network be configured to carry an MTU at least large enough to accommodate the added encapsulation headers. Meanwhile, it is strongly RECOMMENDED that Path MTU Discovery [[RFC1191](#)] [[RFC1981](#)] or Packetization Layer Path MTU Discovery (PLPMTUD) [[RFC4821](#)] is used to prevent or minimize fragmentation. Once an tunnel encapsulator needs to perform fragmentation on an original IP packet before encapsulating, it MUST use the same source UDP port for all fragmented packets so as to ensures these fragmented packets are always forwarded on the same path. Note that fragmentation on the original packets is possible only when the packets are IPv4 packets and the DF bit is not set.

5. Congestion Considerations

[Section 3.1.3 of \[RFC5405\]](#) discussed the congestion implications of UDP tunnels. As discussed in [[RFC5405](#)], because other flows can share the path with one or more UDP tunnels, congestion control [[RFC2914](#)] needs to be considered. As specified in [[RFC5405](#)]:

"IP-based traffic is generally assumed to be congestion- controlled, i.e., it is assumed that the transport protocols generating IP-based traffic at the sender already employ mechanisms that are sufficient to address congestion on the path. Consequently, a tunnel carrying IP-based traffic should already interact appropriately with other traffic sharing the path, and specific congestion control mechanisms for the tunnel are not necessary".

Since IP-in-UDP is only used to carry IP traffic which is generally assumed to be congestion controlled by the transport layer, it generally does not need additional congestion control mechanisms. Furthermore, as it is explicitly stated in the Application Statements ([Section 6](#)), this IP-in-UDP encapsulation method MUST only be used within networks that are well-managed, therefore, congestion control mechanism is not needed.

6. Applicability Statements

This IP-in-UDP encapsulation technology MUST only be used within networks which are well-managed by a service provider and MUST NOT be used within the Internet. In the well-managed network, traffic is well-managed to avoid congestion and fragmentation are not needed.

7. Acknowledgements

Thanks to Vivek Kumar, Carlos Pignataro and Mark Townsley for their valuable comments on the initial idea of this document. Thanks to Andrew G. Malis, Joe Touch and Brian E Carpenter for their valuable comments on this document.

8. IANA Considerations

One UDP destination port number indicating IP needs to be allocated by IANA:

Service Name: IP-in-UDP Transport Protocol(s):UDP
Assignee: IESG <iesg@ietf.org>
Contact: IETF Chair <chair@ietf.org>.
Description: Encapsulate IP packets in UDP tunnels.
Reference: This document.
Port Number: TBD1 -- To be assigned by IANA.

One UDP destination port number indicating IP with DTLS needs to be allocated by IANA:

Service Name: IP-in-UDP-with-DTLS
Transport Protocol(s): UDP
Assignee: IESG <iesg@ietf.org>
Contact: IETF Chair <chair@ietf.org>.
Description: Encapsulate IP packets in UDP tunnels with DTLS.
Reference: This document.
Port Number: TBD2 -- To be assigned by IANA.

9. Security Considerations

The security problems faced with the IP-in-UDP tunnel are exactly the same as those faced with IP-in-IP [[RFC2003](#)] and IP-in-GRE tunnels [[RFC2784](#)]. In other words, the IP-in-UDP tunnel as defined in this document by itself cannot ensure the integrity and privacy of data packets being transported through the IP-in-UDP tunnel and cannot enable the tunnel decapsulator to authenticate the tunnel encapsulator. In the case where any of the above security issues is concerned, the IP-in-UDP tunnel SHOULD be secured with IPsec or DTLS. IPsec was designed as a network security mechanism and therefore it resides at the network layer. As such, if the tunnel is secured with IPsec, the UDP header would not be visible to intermediate routers or switches anymore in either IPsec tunnel or transport mode. As a result, the meaning of adopting the IP-in-UDP tunnel as an alternative to the IP-in-GRE or IP-in-IP tunnel is lost. By comparison, DTLS is better suited for application security and can better preserve network and transport layer protocol information.

Specifically, if DTLS is used, the destination port of the UDP header will be filled with a value (TBD2) indicating IP with DTLS and the source port can still be used as an entropy field for load-sharing purposes.

If the tunnel is not secured with IPsec or DTLS, some other method should be used to ensure that packets are decapsulated and forwarded by the tunnel tail only if those packets were encapsulated by the tunnel head. If the tunnel lies entirely within a single administrative domain, address filtering at the boundaries can be used to ensure that no packet with the IP source address of a tunnel endpoint or with the IP destination address of a tunnel endpoint can enter the domain from outside. However, when the tunnel head and the tunnel tail are not in the same administrative domain, this may become difficult, and filtering based on the destination address can even become impossible if the packets must traverse the public Internet. Sometimes only source address filtering (but not destination address filtering) is done at the boundaries of an administrative domain. If this is the case, the filtering does not provide effective protection at all unless the decapsulator of an IP-in-UDP validates the IP source address of the packet.

10. References

10.1. Normative References

- [RFC0768] Postel, J., "User Datagram Protocol", STD 6, [RFC 768](#), DOI 10.17487/RFC0768, August 1980, <<https://www.rfc-editor.org/info/rfc768>>.
- [RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", [RFC 1191](#), DOI 10.17487/RFC1191, November 1990, <<https://www.rfc-editor.org/info/rfc1191>>.
- [RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", [RFC 1981](#), DOI 10.17487/RFC1981, August 1996, <<https://www.rfc-editor.org/info/rfc1981>>.
- [RFC2003] Perkins, C., "IP Encapsulation within IP", [RFC 2003](#), DOI 10.17487/RFC2003, October 1996, <<https://www.rfc-editor.org/info/rfc2003>>.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.

- [RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", [RFC 2460](#), DOI 10.17487/RFC2460, December 1998, <<https://www.rfc-editor.org/info/rfc2460>>.
- [RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", [RFC 2784](#), DOI 10.17487/RFC2784, March 2000, <<https://www.rfc-editor.org/info/rfc2784>>.
- [RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", [RFC 4301](#), DOI 10.17487/RFC4301, December 2005, <<https://www.rfc-editor.org/info/rfc4301>>.
- [RFC4821] Mathis, M. and J. Heffner, "Packetization Layer Path MTU Discovery", [RFC 4821](#), DOI 10.17487/RFC4821, March 2007, <<https://www.rfc-editor.org/info/rfc4821>>.
- [RFC5405] Eggert, L. and G. Fairhurst, "Unicast UDP Usage Guidelines for Application Designers", [RFC 5405](#), DOI 10.17487/RFC5405, November 2008, <<https://www.rfc-editor.org/info/rfc5405>>.
- [RFC6347] Rescorla, E. and N. Modadugu, "Datagram Transport Layer Security Version 1.2", [RFC 6347](#), DOI 10.17487/RFC6347, January 2012, <<https://www.rfc-editor.org/info/rfc6347>>.
- [RFC6935] Eubanks, M., Chimento, P., and M. Westerlund, "IPv6 and UDP Checksums for Tunneled Packets", [RFC 6935](#), DOI 10.17487/RFC6935, April 2013, <<https://www.rfc-editor.org/info/rfc6935>>.
- [RFC6936] Fairhurst, G. and M. Westerlund, "Applicability Statement for the Use of IPv6 UDP Datagrams with Zero Checksums", [RFC 6936](#), DOI 10.17487/RFC6936, April 2013, <<https://www.rfc-editor.org/info/rfc6936>>.

10.2. Informative References

- [RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", [RFC 2474](#), DOI 10.17487/RFC2474, December 1998, <<https://www.rfc-editor.org/info/rfc2474>>.
- [RFC2914] Floyd, S., "Congestion Control Principles", [BCP 41](#), [RFC 2914](#), DOI 10.17487/RFC2914, September 2000, <<https://www.rfc-editor.org/info/rfc2914>>.

- [RFC3931] Lau, J., Ed., Townsley, M., Ed., and I. Goyret, Ed., "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", [RFC 3931](#), DOI 10.17487/RFC3931, March 2005, <<https://www.rfc-editor.org/info/rfc3931>>.
- [RFC5640] Filsfils, C., Mohapatra, P., and C. Pignataro, "Load-Balancing for Mesh Softwires", [RFC 5640](#), DOI 10.17487/RFC5640, August 2009, <<https://www.rfc-editor.org/info/rfc5640>>.
- [RFC6438] Carpenter, B. and S. Amante, "Using the IPv6 Flow Label for Equal Cost Multipath Routing and Link Aggregation in Tunnels", [RFC 6438](#), DOI 10.17487/RFC6438, November 2011, <<https://www.rfc-editor.org/info/rfc6438>>.
- [RFC6830] Farinacci, D., Fuller, V., Meyer, D., and D. Lewis, "The Locator/ID Separation Protocol (LISP)", [RFC 6830](#), DOI 10.17487/RFC6830, January 2013, <<https://www.rfc-editor.org/info/rfc6830>>.
- [RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", [RFC 7348](#), DOI 10.17487/RFC7348, August 2014, <<https://www.rfc-editor.org/info/rfc7348>>.
- [RFC7510] Xu, X., Sheth, N., Yong, L., Callon, R., and D. Black, "Encapsulating MPLS in UDP", [RFC 7510](#), DOI 10.17487/RFC7510, April 2015, <<https://www.rfc-editor.org/info/rfc7510>>.

Authors' Addresses

Xiaohu Xu
Alibaba Inc

Email: 13910161692@qq.com

Hamid Assarpour
Broadcom

Email: hamid.assarpour@broadcom.com

Shaowen Ma
Mellanox

Email: mashaowen@gmail.com

Daniel Bernier
Canada Bell

Email: daniel.bernier@bell.ca

Darren Dukes
Cisco

Email: ddukes@cisco.com

Shraddha Hegde
Juniper

Email: shraddha@juniper.net

Yiu Lee
Comcast

Email: Yiu_Lee@Cable.Comcast.com

Yongbing Fan
China Telecom

Email: fanyb@gsta.com

