Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: September 9, 2015

A. Wang
China Telecom
S. Jiang
Huawei Technologies Co., Ltd

March 8, 2015

IPv6 Flow Label Reflection
draft-wang-6man-flow-label-reflection-01

Abstract

The current definition of the IPv6 Flow Label focuses mainly on how the packet source forms the value of this field and how the forwarder in-path treats it. In network operations, there are needs to correlate an upstream session and the corresponding downstream session together. This document propose a flow label reflection mechanism that network devices copy the flow label value from received packets to the corresponding flow label field in return packets. This mechanism could simplify the network traffic recognition process in network operations and make the policy for both directions of traffic of one session consistent.

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 http://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 September 9, 2015.

Copyright Notice

Table of Contents

1. Introduction

The IPv6 flow label [RFC6437] in the fixed IPv6 header is designed to differentiate the various flow session of IPv6 traffic; it can accelerate the clarification and treatment of IPv6 traffic by the network devices in its forwarding path. In practice, many current implementations use the 5-tuple {dest addr, source addr, protocol, dest port, source port} as the identifier of network flows. However, transport-layer information, such as the port numbers, is not always in a fixed position, since it follows any IPv6 extension headers that may be present; in contrast, the flow label is at a fixed position in every IPv6 packet and easier to access. In fact, the logic of

finding the transport header is always more complex for IPv6 than for IPv4, due to the absence of an Internet Header Length field in IPv6. Additionally, if packets are fragmented, the flow label will be present in all fragments, but the transport header will only be in one packet. Therefore, within the lifetime of a given transport-

layer connection, the flow label can be a more convenient "handle" than the port number for identifying that particular connection.

The usages of IPv6 flow label, so far as briefly summarized in Section 1.1, only exploit the characteristic of IPv6 flow label in one direction.

In current practice, an application session is often recognized as two separated IP traffics, in two opposite directions. However, from the point view of a service provider, the upstream and downstream of one session should be handled together, particularly, when application-aware operations are placed in the network. A ubiquitous example is that end user initiates a request, with small-scale data transmitted, towards a content server, then the server responds with a large set of follow-up packets. The bi-directional flows should be correlated together and handled with the same policy. Ideally, the request embeds a flow recognition identifier that is accessible and the follow-up response packets carry the same identifier. The flow label is a good choice for the flow recognition identifier.

This document proposes a flow label reflection mechanism so that network devices copy the flow label value from received packets to the corresponding flow label field in return packets. By having the same flow label value in the downstream and upstream of one IPv6 traffic session, the network traffic recognition process and the traffic policy deployment in network operations could be simplified. It may also increase the accuracy of network traffic recognition.

Several applicable scenarios of the IPv6 flow label reflection are also given, in Section 5. For now, this document only considers the scenario in a single administrative domain, although the IPv6 flow label reflection mechanism may also bring benefits into cross domain scenarios.

1.1. Summary of the current usage for IPv6 Flow Label

[RFC6438] describe the usage of IPv6 Flow Label for ECMP and link aggregation in Tunnels; it mainly utilizes the random distribution characteristic of IPv6 flow label. [RFC7098] also describes similar usage in server farms.

All these usage scenarios consider only the usage of IPv6 flow label in one direction, while many bi-directional network traffics need to be treated together.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119] when they appear in ALL CAPS. When these words are not in ALL CAPS (such as "should" or "Should"), they have their usual English meanings, and are not to be interpreted as [RFC2119] key words.

Flow Label Reflection A mechanism/behavior so that a network device copies the value of flow label from a IPv6 flow into a corresponding return IPv6 flow.

Flow Label Reflection Device A network device that applies the flow label reflection mechanism. It is the end of an IPv6 flow and the initiation node of the corresponding return IPv6 flow.

3. Potential Benefit of Flow Label Reflection

With flow label reflection mechanism, the IPv6 Flow Label could be used to correlate the upstream and downstream packets of bi-directional traffics:

- It makes the downstream and upstream of one session be easily recognized. It makes the correlation of traffic and then the recognition of various traffics easier.

- The network operator can easily apply the same policy to the bi-

directional traffic of one interested session

- The traffic analyzer can also easily correlate the upstream and downstream of one session to find the symptoms of various internet protocols.

4. Flow Label Reflection Behaviors on Network Devices

To fulfill the flow label reflection mechanism, the below proposed behaviors on network devices:

- The generation method of IPv6 flow label in source IPv6 node SHOULD follow the guidelines in [RFC6437], that is the IPv6 flow label should be generated randomly and distributed enough.

- On the Flow Label Reflection Device, the value of IPv6 Flow Label from received packets SHOULD be copied into the corresponding flow label field in return packets by the flow label reflection devices.

- The forwarding nodes within the management domain SHOULD follow the specification in [RFC6437], that is the IPv6 flow label SHOULD NOT be modified in the path, unless flow label value in arriving packets is zero. The forwarding nodes MAY follow the specification in [RFC6438] when using the flow label for load balancing by equal cost multipath (ECMP) routing and for link aggregation, particularly for IPv6-in-IPv6 tunneled traffic.

- The network traffic recognition devices, or devices that may have differentiated operations per flow, SHOULD recognize and analyze network traffics based on 3-tuple of {dest addr, source addr, flowlabel}. It SHOULD consider the traffics that have same flow label value and reversed source/dest addr as upstream and downstream of the same flow, match them together to accomplish the traffic recognition process.

- Other network operations MAY also be based on 3-tuple of {dest addr, source addr, flowlabel}.

5. Applicable Scenarios

This section describes some applicable scenarios, which network

operators can benefit from deploying the flow label reflection mechanism. It is not a complete enumeration. More scenarios may be introduced in the future.

5.1. Flow Label Reflection on CP servers

There is rapidly increasing requirement from service providers (SP) to cooperate with the content providers (CP) to provide more accurate services and charging policies based on accurate traffic recognition. The service providers need to recognize the CP/SP's bi-directional traffics at the access edge devices of the network, such as BRAS/PDSN/P-GW devices.

Normally, the burden for these edge devices to recognize the subscriber's upstream traffic is light, because request messages are typically small. But they often need more resource to recognize downstream traffics, which normally contain large data. With flow label reflection on CP servers, recognition based on the 3-tuple of {dest addr, source addr, flowlabel} would reduce the consumption of recognition and make the correlation process much easier.

In this scenario, the CP servers would be the Flow Label Reflection Devices. They copy the flow label value from received upstream user request packets to the corresponding flow label field in return downstream packets.

The access edge devices of service provider scrutinize the subscriber's upstream IPv6 traffic and record the binding of 3-tuple and traffic-specific policy. If the flow label is zero, the access edge devices must rewrite the flow label value according to local policy. With the recorded binding information, the access edge devices can easily recognize and match the downstream packet to the previous recognized upstream packet, by just accessing 3-tuple. The edge devices can then apply the corresponding traffic policy to the upstream/downstream of the session to the cooperated CP.

Note: this mechanism may not reliable when the CP servers are not directly connected to the service provider, because there is no guarantee the flow label would not be changed by intermediate devices in other administrative domains.

5.2. Flow Label Reflection for Bi-direction Tunnels

Tunnel is ubiquitous within service provider networks. It is very difficult (important if the tunnel is encrypted) for intermediate network devices to recognize the inner encapsulated packet, although such recognition could be very helpful in some scenarios, such as traffic statistics, network diagnoses, etc. Furthermore, such recognition normally requires to correlate bi-direction traffic together. The flow label reflection mechanism could provide help in such requirement scenarios.

In this scenario, the tunnel end devices would be the Flow Label Reflection Devices. They record the flow label value from received tunnel packets, and copy it to the corresponding flow label field in return packets, which can be recognized by 5-tuple or 3-tuple of the inner packet at the tunnel end devices.

The tunnel initiating devices should generate different flow label values for different inner flow traffics based on their 5-tuple or 3-tuple in accordance with [RFC6437]. Note: if the inner flow is encryption in ESP model [RFC4303], the transport-layer port numbers are inaccessiable. In such case, 5-tuple is not available.

Then the intermediate network device can easily distinguish the different flow within the same tunnel transport link and correlate bi-direction traffics of same flow together. This can also increase the service provider's traffic control capabilities.

This mechanism can also work when the encapsulated traffics are IPv4 traffics, such as DS-Lite scenario [RFC6333]. The IPv4 5-tuple may be used as the input for the flow label generation.

5.3. Flow Label Reflection on edge devices

If the flow label reflection mechanisms have been applied on peer host, the service provider could always use it for bi-directional traffic recognition. However, there is no guarantee the flow label would not be changed by intermediate devices in other administrative domains. Therefore, to make the flow label value trustful, the edge devices need to validate the flow label reflection.

In this scenario, the edge devices would be the (backup) Flow Label Reflection Devices. They record the flow label value from the packets that leave the domain. When the corresponding flow label field in return packets are recognized by 5-tuple or 3-tuple at the edge devices, the edge devices should check the flow label as below:

- if the flow label matches the record value, it remains;
- if the flow label is zero, the edge devices copy the record value into it;
- if the flow label is non-zero, but does not matches the record value, the edge devices can decide the flow label are modified by other intermediate devices (with the assumption the peer host has reflect the original flow label), then restore the flow label using the record value.

Then the network recognition devices located anywhere within the service provider network could easily correlate bi-directional traffics together, and apply traffic-specific policy accordingly.

5.4. Misc Possible Scenarios

In the below scenarios, the flow label reflection mechanism needs to be combined with other mechanisms in order to achieve the design goals.

5.4.1. Aid to mitigate the ND cache DDoS Attack

Neighbor Discovery Protocol [RFC4861][RFC4861] is vulnerable for the possible DDoS attack to the device's ND cache, see section 11.1, [RFC4861]. There are many proposals are aiming to mitigate this problem, but none of them are prevalent now. It is mainly because that there is no obvious mechanism to assure the validation of the NS/RS packet on the first arrival, the receiving node by default will cache the link-layer address of the NA packet. Reverse detection mechanisms can be added to solve this issue. However, for reverse detection mechanisms, there would be another issue: how to pair the return NA/RA packet with the NS/RS packet on the sending node. It

can be solved by applying the flow label reflection mechanism in the

return NA/RA packet. Then the sending node can pair the reverse detect NS/RS packet with original NA/RA packet and response to the reverse detect NS/RS packet correctly. Only the NS/RS packet that passed the reverse detection validation will be accept by the node and the link-layer address within it will be cached.

5.4.2. Improve the efficiency of PTB problem solution in load-balance environment

[I-D.v6ops-pmtud-ecmp-problem] introduces the Packet Too Big [RFC4443] problem in load-balance environment. The downstream packet from a server, which responses to a client request message, may meet a forwarding node that rejects the packet for "too big" reason. The PTB error ICMPv6 message should be returned to the original server. However, it requests the load balancer to distribute the PTB error ICMPv6 message based on the information of the invoking packet within the ICMPv6 packet, not the ICMPv6 packet itself. The load balancer needs to obtain the source IP address and transport port information within the invoking packet.

However, if both the server and the forwarding node that generates the PTB message apply the flow label reflection mechanism, the PTB error ICMPv6 message would have the same flow label with the original client request message. Then, the load balancer, that follows [RFC7098], could easily forward the PTB packet to same server without parsing the transport port in the invoking packet, thus increases the efficiency.

6. Deployment Consideration

The IPv6 flow label reflection mechanism requires the "Flow Label Reflection Device" to be stateful, store the flow label value and copy it to the corresponding return packet. Such change cannot be accomplished within a short term, and therefore the deployment of this mechanism will be accomplished gradually. During the incremental deployment period, the traditional recognition mechanisms, which are more expensive, would coexist. The traffics that could not be recognized by 3-tuple of {dest addr, source addr, flowlabel} could fall back to the traditional process or be skipped over by advanced services. The more devices support the flow label reflection mechanism, the less consumption for traffic recognition from the network management perspective, or the better coverage of advanced services that are based on the traffic recognition.

7. Security Considerations

Security aspects of the flow label are discussed in [RFC6437]. A malicious source or man-in-the-middle could disturb the traffic recognition by manipulating flow labels. However, the worst case is that fall back to the current practice that an application session is often recognized as two separated IP traffics. The flow label does not significantly alter this situation.

Specifically, the IPv6 flow label specification [RFC6437] states that "stateless classifiers should not use the flow label alone to control load distribution." This is answered by also using the source and destination addresses with flow label.

8. IANA Considerations

This draft does not request any IANA action.

9. Acknowledgements

The authors would like to thanks Brian Carpenter, who gave many useful advices. The authors would also like to thanks the valuable comments made by Fred Baker, Lee Howard, Mark ZZZ Smith, Jeroen Massar, Florent Fourcot and other members of V6OPS WG. Also, special thanks for Florent Fourcot, who have implemented the flow label reflection mechanims in the Linux.

This document was produced using the xml2rfc tool [RFC2629].

10. References

10.1. Normative References

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

[RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.

[RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.

[RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.

[RFC6437] Amante, S., Carpenter, B., Jiang, S., and J. Rajahalme, "IPv6 Flow Label Specification", RFC 6437, November 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.

10.2. Informative References

[I-D.v6ops-pmtud-ecmp-problem] Byerly, M., Hite, M., and J. Jaeggli, "Close encounters of the ICMP type 2 kind (near misses with ICMPv6 PTB)", draft-v6ops-pmtud-ecmp-problem-00 (work in progress), August 2014.

[RFC4303] Kent, S., "IP Encapsulating Security Payload (ESP)", RFC 4303, December 2005.

[RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011.

[RFC7098] Carpenter, B., Jiang, S., and W. Tarreau, "Using the IPv6 Flow Label for Load Balancing in Server Farms", RFC 7098, January 2014.

Authors' Addresses

Aijun Wang
China Telecom
Beijing Research Institute, China Telecom Cooperation Limited
No.118, Xizhimenneidajie, Xicheng District, Beijing 100035
China

Phone: 86-10-58552347

Email: wangaj@ctbri.com.cn

Sheng Jiang
Huawei Technologies Co., Ltd
Q14, Huawei Campus, No.156 Beiqing Road
Hai-Dian District, Beijing, 100095
P.R. China

Email: jiangsheng@huawei.com