Network Working Group
Internet-Draft
Updates: 3697 (if approved)
Intended status: Experimental
Expires: October 15, 2010

B. Carpenter
Univ. of Auckland
S. Jiang
Huawei Technologies Co., Ltd

April 13, 2010

Update to the IPv6 flow label specification
draft-carpenter-6man-flow-update-02

Abstract

Various uses proposed for the IPv6 flow label are incompatible with its existing specification. This document describes changes to the specification that permit additional use cases as well as allowing continued use of the previous specification.

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 October 15, 2010.

Copyright Notice

Table of Contents

1. Introduction 3
2. Normative Notation 4
3. Changes to specification 4
4. Discussion 5
5. Security Considerations 6
6. IANA Considerations 7
7. Acknowledgements 7
8. Change log 7
9. References 7
9.1. Normative References 7
9.2. Informative References 7
Appendix A. Alternative Approaches 9
Authors' Addresses 10

1. Introduction

The flow label field in the IPv6 header is reserved but left experimental by [RFC2460] and is specified by [RFC3697]. We quote three rules from that RFC:

a. "The Flow Label value set by the source MUST be delivered unchanged to the destination node(s)."
b. "IPv6 nodes MUST NOT assume any mathematical or other properties of the Flow Label values assigned by source nodes."
c. "Router performance SHOULD NOT be dependent on the distribution of the Flow Label values. Especially, the Flow Label bits alone make poor material for a hash key."

The second two rules appear to forbid a usage in which the bits of the flow label are encoded with a specific semantic meaning, or are assumed to have any particular property such as randomness. However, both before and after these rules were laid down, a considerable number of proposals for use of the flow label have been published that seem incompatible with them. Examples are [I-D.conta-ipv6-flow-label], [I-D.conta-diffserv-ipv6-fl-classifier], [I-D.chakravorty-6lsa], [I-D.banerjee-flowlabel-ipv6-qos], [I-D.metzler-ipv6-flowlabel], [LeeKim], [LinTseng], and [Prakash]. These authors propose use cases in which some combination of the following options apply:

- The flow label may be changed by intermediate systems.
- It doesn't matter if the flow label is changed, because the receiver doesn't use it.
- Some or all bits of the flow label are coded: they have specific meanings understood by routers and switches along the path.
- The coding is related to the required quality of service, as well as identifying a flow.
- The label is used to control forwarding or switching in some way.

These proposals all require either some form of encoding of semantics in the bits of the flow label, or the ability for routers to modify the flow label, or both. Thus they appear to infringe the rules from

RFC 3697 quoted above.

Although [I-D.roberts-inband-qos-ipv6] does not explicitly consider the flow label, it requests hop-by-hop functionality in IPv6 packets very similar to what is needed by the above proposals.

We can conclude that a considerable number of researchers and designers are stymied by RFC 3697. On the other hand, proposals such as [I-D.martinbeckman-ietf-ipv6-fls-ipv6flowswitching], [I-D.martinbeckman-ietf-ipv6-amp-ipv6hcamp], [I-D.blake-ipv6-flow-label-nonce], and [I-D.carpenter-flow-ecmp] appear to be compatible with RFC 3697. The latter two are based on

the originator of a packet choosing a pseudo-random flow label for each flow. Thus, we can also conclude that there is a useful role for this approach too. The proposal below is intended to resolve this dilemma by allowing both approaches to co-exist.

2. Normative Notation

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].

3. Changes to specification

We note that although RFC 3697 requires the flow label to be delivered unchanged, it is not included in any transport layer pseudo-header checksums nor in IPsec authentication [RFC4302]. We also note that both RFC 2460 and RFC 3697 define the default flow label to be zero. At the time of writing, this is the observed value in an overwhelming proportion of IPv6 packets; neither operating systems nor applications currently set it, and routers do not rely on it. Thus there is no reason to expect operational difficulties if a careful change is made to the rules of RFC 3697.

The purpose of the proposed change is that the flow label should be available for domain-specific use, with locally defined semantics, without preventing uses that are compatible with RFC 3697. There should be no impact on specifications other than RFC 3697 and no

impact on currently operational software and hardware.

The rules of RFC 3697 are modified as follows:

1. If and only if the flow label in an IPv6 packet has the default value of zero, then a router MAY set it to a value between between 1 and 0xFFFFF. This option modifies the rule that the flow label must be delivered unchanged, by allowing exactly one router to set it if the source host did not set it.
2. If this is done, all packets in a given flow MUST be given the same flow label value. A flow is defined in this case as all packets with the same source and destination IPv6 addresses and port numbers and the same transport protocol number, i.e., the same final Next Header value [RFC2460]. This rule constrains the definition of a flow in RFC 3697 for the specific case that a router sets the flow label. However, it does not constrain the bits of the flow label in any particular way.
3. An administratively defined domain containing hosts and routers MAY use a locally defined scheme for the bits of the flow label. This is known as a flow label domain, analogous to a

differentiated services domain [RFC2474]. Hosts in such a domain MUST be configured either to set a default (zero) flow label in all IPv6 packets, or to apply the locally defined scheme. When a locally defined scheme is used, packets entering the flow label domain from outside might contain an invalid label according to that scheme. Therefore, border routers MAY treat all packets entering the flow label domain as if they had a default (zero) flow label. This option will be applied in any case where incorrect flow label formats might cause unpredictable behaviour.

4. Unless a locally defined scheme for the bits of the flow label is in use, the label, whether set by the source host according to RFC 3697, or by a router according to rules 1 and 2 above, SHOULD contain a pseudo-random value between 1 and 0xFFFFF. The intention of this rule is to encourage load balancing solutions based on using the flow label as input to a hash function, e.g., [I-D.carpenter-flow-ecmp], or to enable behaviour such as defined in [I-D.blake-ipv6-flow-label-nonce]. This recommendation constrains the choice of flow label value more than RFC 3697.

The following are the consequences of the above rules combined with those in RFC 3697:

- Sending hosts that are not updated will in practice continue to

send all-zero labels. If there is no locally defined scheme in use along the path taken by a packet, the label will be delivered unchanged.

- Sending hosts wishing to rely only on RFC 3697 behaviour will choose labels between 1 and 0xFFFFF, which should be pseudo-random according to rule 4 above.
- Locally defined behaviour of the flow label will be limited to consistent administratively defined domains.
- Sending hosts wishing to use locally defined behaviour may continue to send all-zero labels, relying on a router in the local flow label domain to set a value according to rules 1, 2 and 3 above. Alternatively, they may set a label according to locally defined rules.
- Routers wishing to implement a locally defined behaviour will set a label according to rules 1, 2 and 3 above, if and only if the incoming flow label is all-zero; if the flow label is not all-zero they must not change it, according to rule 1 above.
- There is an exception to immutability for border routers of a flow label domain, when external packets arrive with a non-zero flow label.

4. Discussion

Hosts that set a default (zero) flow label and ignore the flow label on receipt will be unaffected by implementations of this

specification. In general, it is assumed that hosts will ignore the flow label on receipt; it cannot be safely used as an end-to-end transport or application layer signal of any kind.

Routers that ignore the flow label will be unaffected by implementations of this specification.

Hosts that set a default (zero) flow label and are in a flow label domain where routers adopt rules 1, 2, and 3 or 4 in Section 3 will benefit from whatever flow label handling is used in the local domain. Clearly, the rules b and c quoted from RFC 3697 in Section 1 have no effect within the local domain, where the locally defined rules (whatever they are) replace them.

Hosts and routers that adopt rule 4 by setting a pseudo-random flow

label will enhance the performance of any load balancing devices that include the flow label in the hash used to select a particular path or server, even when packets leave the local domain. Again, rules b and c have no effect.

If a locally defined flow label scheme is used, so that rule 4 is not implemented, then packets leaving the local domain may contain flow label values that are not pseudo-random. However, because of rule 2, the flow label will still be part of the signature of a single packet flow, so it may still be used as part of a load balancing hash. Rules b and c remain true but do not prevent such usage. To allow for these rules, the load balancing hash function needs to be be designed to allow for a possibly non-random flow label, and traffic containing a non-random flow label might not gain full benefit from load balancing.

The rules defined in this document are intended to allow both RFC 3697 usage of the flow label in the general case, and a wide variety of locally defined schemes. Such schemes do not need any global assignments of bits in the flow label, and should not have noticeable impact on backwards compatibility or on domains not using them.

5. Security Considerations

The flow label is not protected in any way and can be forged by an on-path attacker. On the other hand, a pseudo-random flow label cannot be readily guessed by an off-path attacker. See RFC 3697 for further discussion.

6. IANA Considerations

This document requests no action by IANA.

7. Acknowledgements

The authors are grateful to Qinwen Hu for general discussion about

the flow label and for his work in searching the literature. Valuable comments and contributions were made by Remi Despres, Mark Smith and participants in the 6man working group.

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

8. Change log

draft-carpenter-6man-flow-update-02: revised and simplified according to WG discussion, 2010-04-13

draft-carpenter-6man-flow-update-01: revised according to mail list discussion, 2010-03-05

draft-carpenter-6man-flow-update-00: original version, 2010-02-18

9. References

9.1. Normative References

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

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC3697] Rajahalme, J., Conta, A., Carpenter, B., and S. Deering, "IPv6 Flow Label Specification", RFC 3697, March 2004.

9.2. Informative References

[I-D.banerjee-flowlabel-ipv6-qos]
Banerjee, R., "A Modified Specification for use of the IPv6 Flow Label for providing An efficient Quality of Service using hybrid approach", draft-banerjee-flowlabel-ipv6-qos-03 (work in progress), April 2002.

[I-D.blake-ipv6-flow-label-nonce]

Blake, S., "Use of the IPv6 Flow Label as a Transport-Layer Nonce to Defend Against Off-Path Spoofing Attacks", draft-blake-ipv6-flow-label-nonce-02 (work in progress), October 2009.

[I-D.carpenter-flow-ecmp]
Carpenter, B., "Using the IPv6 flow label for equal cost multipath routing in tunnels", draft-carpenter-flow-ecmp-01 (work in progress), February 2010.

[I-D.chakravorty-6lsa]
Chakravorty, S., Bush, J., and J. Bound, "IPv6 Label Switching Architecture", draft-chakravorty-6lsa-03 (work in progress), July 2008.

[I-D.conta-diffserv-ipv6-fl-classifier]
Conta, A. and J. Rajahalme, "Amodel for Diffserv use of the IPv6 Flow Label Specification", draft-conta-diffserv-ipv6-fl-classifier-01 (work in progress), November 2001.

[I-D.conta-ipv6-flow-label]
Conta, A. and B. Carpenter, "A proposal for the IPv6 Flow Label Specification", draft-conta-ipv6-flow-label-02 (work in progress), July 2001.

[I-D.martinbeckman-ietf-ipv6-amp-ipv6hcamp]
Beckman, M., "IPv6 Header Compression via Addressing Mitigation Protocol (IPv6 AMP)", draft-martinbeckman-ietf-ipv6-amp-ipv6hcamp-01 (work in progress), March 2007.

[I-D.martinbeckman-ietf-ipv6-fls-ipv6flowswitching]
Beckman, M., "IPv6 Dynamic Flow Label Switching (FLS)", draft-martinbeckman-ietf-ipv6-fls-ipv6flowswitching-03 (work in progress), March 2007.

[I-D.metzler-ipv6-flowlabel]
Metzler, J. and S. Hauth, "An end-to-end usage of the IPv6 flow label", draft-metzler-ipv6-flowlabel-00 (work in progress), November 2000.

[I-D.roberts-inband-qos-ipv6]
Roberts, L. and J. Harford, "In-Band QoS Signaling for IPv6", draft-roberts-inband-qos-ipv6-00 (work in progress), July 2005.

[LeeKim] Lee, I. and S. Kim, "A QoS Improvement Scheme for Real-Time Traffic Using IPv6 Flow Labels", Lecture Notes in Computer Science Vol. 3043, 2004.

[LinTseng] Lin, C., Tseng, P., and W. Hwang, "End-to-End QoS Provisioning by Flow Label in IPv6", JCIS , 2006.

[Prakash] Prakash, B., "Using the 20 bit flow label field in the IPv6 header to indicate desirable quality of service on the internet", University of Colorado (M.Sc. Thesis), 2004.

[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, December 1998.

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

[RFC4302] Kent, S., "IP Authentication Header", RFC 4302, December 2005.

Appendix A. Alternative Approaches

Two more complex alternative approaches were considered and rejected.

The first was to distinguish locally significant flow labels from those conforming to RFC 3697 by setting or clearing the most significant bit (MSB) of the flow label. This led to quite complicated rules and was not considered practical.

The second was to use a specific differentiated services code point (DSCP)[RFC2474] in the Traffic Class octet instead of the MSB of the flow label itself, to flag a locally defined behaviour. A more elaborate version of this was proposed in [I-D.martinbeckman-ietf-ipv6-fls-ipv6flowswitching]. There are two issues with this approach. One is that DSCP values are themselves only locally significant, inconsistent with the end-to-end nature of the original flow label definition. Secondly, it seems unwise to meld the semantics of differentiated services, which are currently deployed, with the unknown future semantics of flow label usage. However, this approach, while not recommended, does not appear to violate any basic principles if applied strictly within a single

differentiated services domain.

Authors' Addresses

Brian Carpenter
Department of Computer Science
University of Auckland
PB 92019
Auckland, 1142
New Zealand

Email: brian.e.carpenter@gmail.com

Sheng Jiang
Huawei Technologies Co., Ltd
KuiKe Building, No.9 Xinxi Rd.,
Shang-Di Information Industry Base, Hai-Dian District, Beijing
P.R. China

Email: shengjiang@huawei.com