6MAN
Internet-Draft
Intended status: Informational
Expires: August 28, 2011

S. Amante
Level 3
B. Carpenter
Univ. of Auckland
S. Jiang
Huawei Technologies Co., Ltd

February 26, 2011

Rationale for update to the IPv6 flow label specification draft-ietf-6man-flow-update-03

Abstract

Various published proposals for use of the IPv6 flow label are incompatible with its original specification in RFC 3697. Furthermore, very little practical use is made of the flow label, partly due to some uncertainties about the correct interpretation of the specification. This document discusses and motivates changes to the specification in order to clarify it, and to introduce some additional flexibility.

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 August 28, 2011.

Copyright Notice

Copyright (c) 2011 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 (http://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 3
2. Impact of current specification 3
3. Changes to specification 6
4. Discussion 7
5. Security Considerations 8
6. IANA Considerations 8
7. Acknowledgements 8
8. Change log 9
9. Informative References 9
Appendix A. Alternative Approaches 10
Authors' Addresses 11

1. Introduction

The flow label field in the IPv6 header was reserved but left experimental by [RFC2460], which mandates only that "Hosts or routers that do not support the functions of the Flow Label field are required to set the field to zero when originating a packet, pass the field on unchanged when forwarding a packet, and ignore the field when receiving a packet."

The flow label field was normatively specified by [RFC3697]. In particular, 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."

Additionally, RFC 3697 leaves it undefined what method a host should adopt by default to choose the value of the flow label, if no specific method is in use. It was expected that various signaling methods might be defined for agreeing on values of the flow label, but no such methods have been standardised, except a pre-existing option in RSVP [RFC2205].

The flow label is hardly used in practice in widespread IPv6 implementations, although some operating systems do set it [McGann05]. To some extent this is due to the main focus being on basic deployment of IPv6, but the absence of a default method of choosing the flow label value means that most host implementations simply set it to zero. There is also anecdotal evidence that the rules quoted above have led to uncertainty about exactly what is possible. Furthermore, various use cases have been proposed that infringe one or another of the rules. None of these proposals has been accepted as a standard and in practice there is no significant deployment of any mechanism to set the flow label.

The intention of this document is to explain this situation in more detail and to motivate changes to RFC 3697 intended to remove the uncertainties and encourage active usage of the flow label. It does not formally update RFC 3697, but it serves as background material for [I-D.ietf-6man-flow-3697bis].

2. Impact of current specification

Rule (a) makes it impossible for the routing system to use the flow

label as any form of dynamic routing tag. This was a conscious choice in the early design of IPv6 and there appears to be no practical possibility of revisiting this choice at this stage in the deployment of IPv6, which uses conventional routing mechanisms like those used for IPv4. However, this rule also makes it impossible to make any use at all of the flow label unless hosts choose to set it. It also forbids clearing the flow label for security reasons.

This last point highlights the security properties, or rather the lack of them, of the flow label. The flow label field is always unprotected as it travels through the network, because there is no IPv6 header checksum, and the flow label is not included in transport pseudo-header checksums, nor in IPsec checksums. As a result, intentional and malicious changes to its value cannot be detected. Also, it could be used as a covert data channel, since apparently pseudo-random flow label values could in fact consist of covert data. If the flow label were to carry quality of service semantics, then like the diffserv code point [RFC2474], it would not be intrinsically trustworthy across domain boundaries. As a result, some security specialists believe that flow labels should be cleared for safety. These points must be considered when discussing the immutability of the flow label across domain boundaries.

Rule (b) appears to forbid any usage in which the bits of the flow label are encoded with a specific semantic meaning. However, the words "MUST NOT assume" are to be interpreted precisely - if a router knows by configuration or by signaling that the flow label has been assigned in a certain way, it can make use of that knowledge. It is not made clear by the rule that there is an implied distinction between stateless models (in which case no assumption may be made) and stateful models (in which the router has explicit knowledge).

If the word "alone" is overlooked, rule (c) has sometimes been interpreted to forbid the use of the flow label as part of a hash used by load distribution mechanisms. In this case too, the word "alone" is to be interpreted precisely - a router is allowed to combine the flow label value with other data in order to produce a uniformly distributed hash.

Both before and after these rules were laid down, a considerable number of proposals for use of the flow label were published that seem incompatible with them. Numerous examples and an analysis are presented in [I-D.hu-flow-label-cases]. Those examples propose use cases in which some or all of the following 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 encoded: they have specific meanings understood by routers and switches along the path.
- The encoding is related to the required quality of service, as well as identifying a flow.
- The flow 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.

We can conclude that a considerable number of researchers and designers have been stymied by RFC 3697. On the other hand, some other proposals discussed in [I-D.hu-flow-label-cases] appear to be compatible with RFC 3697. Several are based on the originator of a packet choosing a pseudo-random flow label for each flow, which is one option suggested in RFC 3697. Thus, we can also conclude that there is a useful role for this approach.

If our goal is for the flow label to be used in practice, the conflict between the various approaches creates a dilemma. There appear to be two major options:

1. Discourage locally defined and/or stateful use of the flow label. Strengthen RFC 3697 to say that hosts SHOULD set a pseudo-random label value, without creating state, which would clarify and limit its possible uses. In particular, its use for load distribution and balancing would be encouraged.
2. Relax the rules to encourage locally defined and/or stateful use of the flow label. This approach would make the flow label completely mutable and would exclude use cases depending on strict end-to-end immutability. It would encourage applications of a pseudo-random flow label, such as load distribution, on a local basis, but it would exclude end-to-end applications.

During 2010 there was considerable debate about these options and variants of them, with a variety of proposals in previous versions of this document and in mailing list discussions. After these discussions, there appears to be a view that simplicity should prevail, and that complicated proposals such as defining quality of service semantics in the flow label, or sub-dividing the flow label field into smaller sub-fields, will not prove efficient or deployable, especially in high speed routers. There is also a clearly expressed view that using the flow label for various forms of stateless load distribution is the best simple application for it. At the same time, it is necessary to recognize that the strict immutability rule has drawbacks as noted above.

Even under the rules of RFC 3697, the flow label is intrinsically untrustworthy, because modifications en route cannot be detected. For this reason, even with the current strict immutability rule, downstream nodes cannot rely on the value being unchanged. In this sense, any use of the flow label must be viewed as an optimisation on a best effort basis; a packet with a changed (or zero) flow label value should never cause a hard failure.

The remainder of this document discusses specific modifications to the standard, which are defined normatively in a companion document [I-D.ietf-6man-flow-3697bis].

3. Changes to specification

Although RFC 3697 requires the flow label to be delivered unchanged, as noted above, it is not included in any transport layer pseudo-header checksums nor in IPsec authentication [RFC4302]. 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; the most widespread operating systems and applications do not 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.

In particular, the facts that the label is not checksummed and rarely used mean that the current strict immutability of the label can be moderated without serious operational consequences.

The purposes of the proposed changes are to remove the uncertainties left by RFC 3697, in order to encourage setting of the flow label by default, and to enable its generic use. The proposed generic use is to encourage pseudo-random flow labels that can be used to assist load distribution balancing. There should be no impact on existing IETF specifications other than RFC 3697 and no impact on currently operational software and hardware.

A secondary purpose is to modify the immutability of the flow label in a limited way, to allow hosts that do not set the flow label to benefit from it nevertheless. The fact that the flow label may in practice be changed en route is also reflected in the reformulation of the rules.

A general description of the changes follows. The normative text is to be found in [I-D.ietf-6man-flow-3697bis].

The definition of a flow is subtly changed from RFC 3697 to allow any node, not just the source node, to set the flow label value.

However, it is recommended that sources should set a pseudo-random flow label value in all flows, replacing the less precise recommendation made in Section 3 of RFC 3697. Both stateful and stateless methods of assigning a pseudo-random value could be used.

Section 3 of RFC 3697 also allows nodes to participate in an unspecified method of flow state establishment. The changes do not remove that option, but it is made clear that stateless models are also possible and are the recommended default.

The main novelty is that a forwarding node (typically a first-hop or ingress router) may set the flow label value if the source has not done so, according to the same recommendations that apply to the source. This might place a considerable processing load on ingress routers, even if they adopted a stateless method of flow identification and label assignment.

The immutability of the flow label, once it has been set, is not changed. However, some qualifications are placed on this property, to allow for the fact that the flow label is an unprotected field and might be changed undetectably. No Internet-wide mechanism can depend mathematically on immutable flow labels. The new rules require that flow labels exported to the Internet should always be either zero or pseudo-random, but even this cannot be relied on mathematically. Use cases need to be robust against non-conforming flow label values. This will also enhance compatibility with any legacy hosts that set the flow label according to RFC 2460 or RFC 3697.

4. Discussion

The following are some practical consequences of the above changes:

- Sending hosts that are not updated will in practice continue to send all-zero labels. If there is no label-setting router along the path taken by a packet, the label will be delivered as zero.
- Sending hosts conforming to the new specification will by default choose pseudo-random labels between 1 and 0xFFFFF.
- Sending hosts may continue to send all-zero labels, in which case an ingress router may set pseudo-random labels between 1 and 0xFFFFF.
- The flow label is no longer unrealistically asserted to be strictly immutable; it is recognised that it may, incorrectly, be changed en route. In some circumstances this will break end-to-end usage, e.g. potential detection of third-party spoofing attacks [I-D.gont-6man-flowlabel-security].
- The expected default usage of the flow label is some form of stateless load distribution, such as the ECMP/LAG usage defined in [I-D.carpenter-flow-ecmp].

- If the new rules are followed, all IPv6 traffic flows on the Internet should have zero or pseudo-random flow label values.

From an operational viewpoint, existing IPv6 hosts that set a default (zero) flow label value and ignore the flow label on receipt will be unaffected by implementations of the new specification. In general, it is assumed that hosts will ignore the value of the flow label on receipt; it cannot be relied on as an end-to-end signal. However, this doesn't apply if a cryptographically generated label is being used to detect attackers [I-D.gont-6man-flowlabel-security].

Similarly, routers that ignore the flow label will be unaffected by implementations of the specification.

Hosts that set a default (zero) flow label but are in a domain where routers set a pseudo-random label as recommended in Section 3 will benefit from whatever flow label handling is used on the path.

Hosts and routers that adopt the recommended pseudo-random mechanism 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.

5. Security Considerations

See [I-D.ietf-6man-flow-3697bis] and [I-D.gont-6man-flowlabel-security] for full 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 Fred Baker, Steve Blake, Remi Despres, Alan Ford, Fernando Gont, Brian Haberman, Tony Hain, Joel Halpern, Chris Morrow, Thomas Narten, Pekka Savola, Mark Smith, Pascal Thubert, Iljitsch van Beijnum, and other participants in the 6man working group.

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

8. Change log

draft-ietf-6man-flow-update-03: updated to be in styep with RFC 3697bis, 2011-02-26

draft-ietf-6man-flow-update-02: repurposed as rationale for update of RFC 3697, 2011-01-31

draft-ietf-6man-flow-update-01: clarified that this is not a formal update of RFC 3697, clarified text about domains exporting inappropriate labels, 2011-01-10

draft-ietf-6man-flow-update-00: adopted as WG document at IETF 79, mutability rules adjusted according to WG discussion, 2010-12-03

draft-carpenter-6man-flow-update-04: even more simplified according to WG discussion, 2010-09-16

draft-carpenter-6man-flow-update-03: futher simplified according to WG discussion, 2010-05-07

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. Informative References

[I-D.carpenter-flow-ecmp]
Carpenter, B. and S. Amante, "Using the IPv6 flow label for equal cost multipath routing and link aggregation in tunnels", draft-carpenter-flow-ecmp-03 (work in progress), October 2010.

[I-D.gont-6man-flowlabel-security]
Gont, F., "Security Assessment of the IPv6 Flow Label", draft-gont-6man-flowlabel-security-01 (work in progress), November 2010.

[I-D.hu-flow-label-cases]
Hu, Q. and B. Carpenter, "Survey of proposed use cases for the IPv6 flow label", draft-hu-flow-label-cases-02 (work in progress), September 2010.

[I-D.ietf-6man-flow-3697bis]
Amante, S., Carpenter, B., Jiang, S., and J. Rajahalme, "IPv6 Flow Label Specification", draft-ietf-6man-flow-3697bis-00 (work in progress), January 2011.

[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.

[McGann05]
McGann, O. and D. Malone, "Flow Label Filtering Feasibility", European Conference on Computer Network Defence , 2005.

[RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

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

[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.

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

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

Appendix A. Alternative Approaches

A model was discussed in an earlier version of this document which defined a notion of 'flow label domain' analogous to a differentiated services domain [RFC2474]. This model would have encouraged local usage of the flow label as an alternative to any form of generic use, but it required complex rules for the behaviour of domain boundary routers, and proved controversial in discussion.

Two even more complex alternative approaches were also 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, seems impossible to make fully self-consistent, 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

Shane Amante
Level 3 Communications, LLC
1025 Eldorado Blvd
Broomfield, CO 80021
USA

Email: shane@level3.net

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
Huawei Building, No.3 Xinxi Rd.,
Shang-Di Information Industry Base, Hai-Dian District, Beijing
P.R. China

Email: shengjiang@huawei.com