6MAN
Internet-Draft
Obsoletes: 3697 (if approved)
Updates: 2205, 2460 (if approved)
Intended status: Standards Track
Expires: August 30, 2011

S. Amante
Level 3
B. Carpenter
Univ. of Auckland
S. Jiang
Huawei Technologies Co., Ltd
J. Rajahalme
Nokia-Siemens Networks

February 26, 2011

IPv6 Flow Label Specification
draft-ietf-6man-flow-3697bis-01

Abstract

This document specifies the IPv6 Flow Label field and the minimum requirements for IPv6 nodes labeling flows, IPv6 nodes forwarding labeled packets, and flow state establishment methods. Even when mentioned as examples of possible uses of the flow labeling, more detailed requirements for specific use cases are out of scope for this document.

The usage of the Flow Label field enables efficient IPv6 flow classification based only on IPv6 main header fields in fixed positions.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

A flow is a sequence of packets sent from a particular source to a particular unicast, anycast, or multicast destination that a node desires to label as a flow. A flow could consist of all packets in a specific transport connection or a media stream. However, a flow is not necessarily 1:1 mapped to a transport connection.

Traditionally, flow classifiers have been based on the 5-tuple of the source and destination addresses, ports, and the transport protocol type. However, some of these fields may be unavailable due to either fragmentation or encryption, or locating them past a chain of IPv6 extension headers may be inefficient. Additionally, if classifiers depend only on IP layer headers, later introduction of alternative transport layer protocols will be easier.

The usage of the 3-tuple of the Flow Label and the Source and Destination Address fields enables efficient IPv6 flow classification, where only IPv6 main header fields in fixed positions are used.

The flow label could be used in both stateless and stateful scenarios. A stateless scenario is one where a node that sets the flow label value for all packets of a given flow does not need to store any information about the flow, and any node that processes the flow label in any way also does not need to store any information after a packet has been processed. A stateful scenario is one where a node that sets or processes the flow label value needs to store information about the flow, including the flow label value. A stateful scenario might also require a signaling mechanism to establish flow state in the network.

The flow label can be used most simply in stateless scenarios. This specification concentrates on the stateless model and how it can be used as a default mechanism. Details of stateful models, signaling, specific flow state establishment methods and their related service models are out of scope for this specification. Generic requirements enabling co-existence of different models are set forth in Section 4. The associated scaling characteristics (such as nodes involved in state establishment, amount of state maintained by them, and state growth function) will be specific to particular service models.

The minimum level of IPv6 flow support consists of labeling the flows. A specific goal is to enable and encourage the use of the flow label for various forms of stateless load distribution, especially across Equal Cost Multi-Path (EMCP) and/or Link Aggregation Group (LAG) paths. ECMP and LAG are methods to bond together multiple physical links used to procure the required

capacity necessary to carry an offered load greater than the bandwidth of an individual physical link. IPv6 source nodes SHOULD be able to label known flows (e.g., TCP connections, application streams), even if the node itself does not require any flow-specific treatment. Node requirements for stateless flow labeling are given in Section 3.

This document replaces [RFC3697] and Appendix A of [RFC2460]. A rationale for the changes made is documented in [I-D.ietf-6man-flow-update]. The present document also includes a correction to [RFC2205] concerning the flow label.

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].

2. IPv6 Flow Label Specification

The 20-bit Flow Label field in the IPv6 header [RFC2460] is used by a node to label packets of a flow. A Flow Label of zero is used to indicate packets not part of any flow. Packet classifiers can use the triplet of Flow Label, Source Address, and Destination Address fields to identify which flow a particular packet belongs to. Packets are processed in a flow-specific manner by nodes that are able to do so in a stateless manner, or that have been set up with flow-specific state. The nature of the specific treatment and the methods for flow state establishment are out of scope for this specification. However, any node that sets flow label values according to a stateful scheme MUST ensure that packets conform to Section 3 of the present specification if they are sent outside the network domain using the stateful scheme.

As specified below in Section 3, the normal expectation is that flow label values are uniformly distributed. In this specification, it is recommended below that a pseudo-random method should be used to achieve such a uniform distribution. Intentionally, there are no precise mathematical requirements placed on the distribution or the pseudo-random method.

Once set to a non-zero value, the Flow Label MUST be delivered unchanged to the destination node(s). A forwarding node MUST NOT change the flow label value in an arriving packet if it is non-zero. However, there are two qualifications to this rule:

1. Implementers are advised that forwarding nodes, especially those acting as domain border devices, might nevertheless be configured to change the flow label value in packets. This is undetectable, unless some future version of IPsec authentication [RFC4302]

protects the flow label value.

2. To enable stateless load distribution at any point in the Internet, a network domain should never export packets originating within the domain whose flow label values do not conform to Section 3. However, neither domain border egress routers nor intermediate routers/devices (using a flow-label, for example, as a part of an input-key for a load-distribution hash) can determine by inspection that a value is not part of a uniform distribution. Therefore, if nodes within a domain ignore the recommendations of Section 3, and such packets are forwarded outside the domain, this might result in undesirable operational implications (e.g., congestion, reordering) for not only the inappropriately flow-labelled packets, but also well-behaved flow-labelled packets, during forwarding at various intermediate devices. Thus, a domain must protect its peers by never exporting inappropriately labelled packets originating within the domain. This is why nodes using a stateful scheme must not set the flow label to a non-zero and non-uniformly distributed value if the packet will leave their domain. If it is known to a border router that flow labels originated within the domain are not uniformly distributed, it will need to set outgoing flow labels in the same manner as described for forwarding nodes in Section 3.

There is no way to verify whether a flow label has been modified en route or whether it belongs to a uniform distribution. Therefore, no Internet-wide mechanism can depend mathematically on immutable and uniformly distributed flow labels; they have a "best effort" quality. This leads to the following formal rules:

- Implementers should be aware that the flow label is an unprotected field that could have been accidentally or intentionally changed en route. Implementations MUST take appropriate steps to protect themselves from being vulnerable to denial of service and other types of attack that could result (see Section 6.1).
- Forwarding nodes such as routers and load balancers MUST NOT depend only on Flow Label values being uniformly distributed. In any usage such as a hash key for load distribution, the Flow Label bits MUST be combined at least with bits from other sources within the packet, so as to produce a constant hash value for each flow and a suitable distribution of hash values across flows.

Although uniformly distributed flow label values are recommended below, and will always be helpful for load balancing, it is unsafe to assume their presence in the general case, and the use case needs to work even if the flow label value is zero.

The use of the Flow Label field does not necessarily signal any requirement on packet reordering. Especially, the zero label does

not imply that significant reordering is acceptable.

An IPv6 node that does not set the flow label to a non-zero value, or make use of it in any way, MUST ignore it when receiving or forwarding a packet.

3. Stateless Flow Labeling Requirements

This section defines the minimum requirements for stateless methods of setting the flow label value.

To enable Flow Label based classification, source nodes SHOULD assign each unrelated transport connection and application data stream to a new flow. A typical definition of a flow for this purpose is any set of packets carrying the same 5-tuple {dest addr, source addr, protocol, dest port, source port}.

It is desirable that flow label values should be uniformly distributed to assist load distribution. It is therefore RECOMMENDED that source hosts support the flow label by setting the flow label field for all packets of a given flow to the same uniformly distributed pseudo-random value. Both stateful and stateless methods of assigning a pseudo-random value could be used, but it is outside the scope of this specification to mandate an algorithm. In a stateless mechanism, the algorithm SHOULD ensure that the resulting flow label values are unique with high probability.

An OPTIONAL algorithm for generating such a pseudo-random value is described in [I-D.gont-6man-flowlabel-security].

[[NOTE TO RFC EDITOR: The preceding sentence should be deleted, and the reference should be changed to Informative, if the cited draft is not on the standards track at the time of publication.]]

A source node which does not otherwise set the flow label MUST set its value to zero.

A node that forwards a flow whose flow label value in arriving packets is zero MAY set the flow label value. In that case, it is RECOMMENDED that the forwarding node sets the flow label field for a flow to a uniformly distributed pseudo-random value.

- The same considerations apply as to source hosts setting the flow label; in particular, the normal case is that a flow is defined by the 5-tuple.
- This option, if implemented, would presumably be used by first-hop or ingress routers. It might place a considerable per-packet processing load on them, even if they adopted a stateless method

of flow identification and label assignment. This is why the principal recommendation is that the source host should set the label.

The preceding rules taken together allow a given network domain to include routers that set flow labels on behalf of hosts that do not do so. They also recommend that flow labels exported to the Internet are always either zero or uniformly distributed.

4. Flow State Establishment Requirements

This section defines the minimum requirements for stateful methods of setting the flow label value.

The node that sets the flow label MAY also take part in flow state establishment methods that result in assigning specific treatments to specific flows, possibly including signaling.

- In this case, unlike the stateless case, a source node MUST ensure that it does not unintentionally reuse Flow Label values it is currently using or has recently used when creating new flows. Flow Label values previously used with a specific pair of source and destination addresses MUST NOT be assigned to new flows with the same address pair within 120 seconds of the termination of the previous flow.
- To avoid accidental Flow Label value reuse, the source node SHOULD select new Flow Label values in a well-defined way and use an initial value that avoids reuse of recently used Flow Label values each time the system restarts. The initial value SHOULD be derived from a previous value stored in non-volatile memory, or in the absence of such history, a randomly generated initial value using techniques that produce good randomness properties SHOULD be used.

To enable stateful flow-specific treatment, flow state needs to be established on all or a subset of the IPv6 nodes on the path from the source to the destination(s). The methods for the state establishment, as well as the models for flow-specific treatment will be defined in separate specifications.

In stateful mechanisms, nodes keeping dynamic flow state MUST NOT assume packets arriving 120 seconds or more after the previous packet of a flow still belong to the same flow, unless a flow state establishment method in use defines a longer flow state lifetime or the flow state has been explicitly refreshed within the lifetime duration.

To enable co-existence of different methods in IPv6 nodes, the

methods MUST meet the following basic requirements:

- o The method MUST provide the means for flow state clean-up from the IPv6 nodes providing the flow-specific treatment. Signaling based methods where the source node is involved are free to specify flow state lifetimes longer than the default 120 seconds.
- o Flow state establishment methods MUST be able to recover from the case where the requested flow state cannot be supported.

5. Essential correction to RFC 2205

[RFC2460] reduced the size of the flow label field from 24 to 20 bits. The references to a 24 bit flow label field on pages 87 and 88 of [RFC2205] are updated accordingly.

6. Security Considerations

This section considers security issues raised by the use of the Flow Label, primarily the potential for denial-of-service attacks, and the related potential for theft of service by unauthorized traffic (Section 6.1). Section 6.2 addresses the use of the Flow Label in the presence of IPsec including its interaction with IPsec tunnel mode and other tunneling protocols. We also note that inspection of unencrypted Flow Labels may allow some forms of traffic analysis by revealing some structure of the underlying communications. Even if the flow label were encrypted, its presence as a constant value in a fixed position might assist traffic analysis and cryptoanalysis.

The flow label is not protected in any way and can be forged by an on-path attacker. On the other hand, a uniformly distributed pseudo-random flow label cannot be readily guessed by an off-path attacker; see [I-D.gont-6man-flowlabel-security] for further discussion.

6.1. Theft and Denial of Service

Since the mapping of network traffic to flow-specific treatment is triggered by the IP addresses and Flow Label value of the IPv6 header, an adversary may be able to obtain better service by modifying the IPv6 header or by injecting packets with false addresses and/or labels. Taken to its limits, such theft-of-service becomes a denial-of-service attack when the modified or injected traffic depletes the resources available to forward it and other traffic streams. A curiosity is that if a DoS attack were undertaken against a given Flow Label (or set of Flow Labels), then traffic containing an affected Flow Label might well experience worse-than-best-effort network performance.

Note that since the treatment of IP headers by nodes is typically unverified, there is no guarantee that flow labels sent by a node are set according to the recommendations in this document. Therefore, any assumptions made by the network about header fields such as flow labels should be limited to the extent that the upstream nodes are explicitly trusted.

Since flows are identified by the 3-tuple of the Flow Label and the Source and Destination Address, the risk of theft or denial of service introduced by the Flow Label is closely related to the risk of theft or denial of service by address spoofing. An adversary who is in a position to forge an address is also likely to be able to forge a label, and vice versa.

There are two issues with different properties: Spoofing of the Flow Label only, and spoofing of the whole 3-tuple, including Source and Destination Address.

The former can be done inside a node which is using or transmitting the correct source address. The ability to spoof a Flow Label typically implies being in a position to also forge an address, but in many cases, spoofing an address may not be interesting to the spoofer, especially if the spoofer's goal is theft of service, rather than denial of service.

The latter can be done by a host which is not subject to ingress filtering [RFC2827] or by an intermediate router. Due to its properties, such is typically useful only for denial of service. In the absence of ingress filtering, almost any third party could instigate such an attack.

In the presence of ingress filtering, forging a non-zero Flow Label on packets that originated with a zero label, or modifying or clearing a label, could only occur if an intermediate system such as a router was compromised, or through some other form of man-in-the-middle attack. However, the risk is limited to traffic receiving better or worse quality of service than intended. For example, if Flow Labels are altered or cleared at random, flow classification will no longer happen as intended, and the altered packets will receive default treatment. If a complete 3-tuple is forged, the altered packets will be classified into the forged flow and will receive the corresponding quality of service; this will create a denial of service attack subtly different from one where only the addresses are forged. Because it is limited to a single flow definition, e.g., to a limited amount of bandwidth, such an attack will be more specific and at a finer granularity than a normal address-spoofing attack.

Since flows are identified by the complete 3-tuple, ingress filtering [RFC2827] will, as noted above, mitigate part of the risk. If the source address of a packet is validated by ingress filtering, there can be a degree of trust that the packet has not transited a compromised router, to the extent that ISP infrastructure may be trusted. However, this gives no assurance that another form of man-in-the-middle attack has not occurred.

A man-in-the-middle denial of service attack specifically directed at flow label handling would involve setting unusual flow labels. For example, an attacker could set all flow labels reaching a given router to the same arbitrary non-zero value, or could perform rapid cycling of flow label values such that the packets of a given flow will each have a different value. Either of these attacks would cause a stateless load distribution algorithm to perform badly and would cause a stateful mechanism to behave incorrectly. For this reason, stateless mechanisms should not use the flow label alone to control load distribution, and stateful mechanisms should include explicit methods to detect and ignore suspect flow label values.

6.2. IPsec and Tunneling Interactions

The IPsec protocol, as defined in [RFC4301], [RFC4302], [RFC4303] does not include the IPv6 header's Flow Label in any of its cryptographic calculations (in the case of tunnel mode, it is the outer IPv6 header's Flow Label that is not included). Hence modification of the Flow Label by a network node has no effect on IPsec end-to-end security, because it cannot cause any IPsec integrity check to fail. As a consequence, IPsec does not provide any defense against an adversary's modification of the Flow Label (i.e., a man-in-the-middle attack).

IPsec tunnel mode provides security for the encapsulated IP header's Flow Label. A tunnel mode IPsec packet contains two IP headers: an outer header supplied by the tunnel ingress node and an encapsulated inner header supplied by the original source of the packet. When an IPsec tunnel is passing through nodes performing flow classification, the intermediate network nodes operate on the Flow Label in the outer header. At the tunnel egress node, IPsec processing includes removing the outer header and forwarding the packet (if required) using the inner header. The IPsec protocol requires that the inner header's Flow Label not be changed by this decapsulation processing to ensure that modifications to label cannot be used to launch theft- or denial-of-service attacks across an IPsec tunnel endpoint. This document makes no change to that requirement; indeed it forbids changes to the Flow Label.

When IPsec tunnel egress decapsulation processing includes a

sufficiently strong cryptographic integrity check of the encapsulated packet (where sufficiency is determined by local security policy), the tunnel egress node can safely assume that the Flow Label in the inner header has the same value as it had at the tunnel ingress node.

This analysis and its implications apply to any tunneling protocol that performs integrity checks. Of course, any Flow Label set in an encapsulating IPv6 header is subject to the risks described in the previous section.

6.3. Security Filtering Interactions

The Flow Label does nothing to eliminate the need for packet filtering based on headers past the IP header, if such filtering is deemed necessary for security reasons on nodes such as firewalls or filtering routers.

However, security devices that clear or rewrite non-zero flow label values would be in violation of this specification.

7. IANA Considerations

This document requests no action by IANA.

8. Acknowledgements

Steve Deering and Alex Conta were co-authors of RFC 3697, on which this document is based.

Valuable comments and contributions were made by Fred Baker, Steve Blake, Remi Despres, Alan Ford, Fernando Gont, Brian Haberman, Tony Hain, Joel Halpern, Qinwen Hu, Chris Morrow, Thomas Narten, Mark Smith, Pascal Thubert, Iljitsch van Beijnum, and other participants in the 6man working group.

Contributors to the development of RFC 3697 included Ran Atkinson, Steve Blake, Jim Bound, Francis Dupont, Robert Elz, Tony Hain, Robert Hancock, Bob Hinden, Christian Huitema, Frank Kastenholz, Thomas Narten, Charles Perkins, Pekka Savola, Hesham Soliman, Michael Thomas, Margaret Wasserman, and Alex Zinin.

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

9. Change log

draft-ietf-6man-flow-3697bis-01: update after resolving 11 initial issues, 2011-02-26

draft-ietf-6man-flow-3697bis-00: original version, built from RFC3697 and draft-ietf-6man-flow-update-01, 2011-01-31

10. References

10.1. Normative References

[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.

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

[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.

10.2. Informative References

[I-D.ietf-6man-flow-update]
Amante, S., Carpenter, B., and S. Jiang, "Rationale for update to the IPv6 flow label specification", draft-ietf-6man-flow-update-02 (work in progress), January 2011.

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

[RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, May 2000.

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

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.

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

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

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

Jarno Rajahalme
Nokia-Siemens Networks
TBD
TBD
Finland

Email: jarno.rajahalme@nsn.com