IPv6 Working Group
INTERNET-DRAFT
<draft-rajahalme-ipv6-flow-label-00.txt>

J. Rajahalme
Nokia
A. Conta
Transwitch

Expires: May 2002

November 2001

An IPv6 Flow Label Specification Proposal
draft-rajahalme-ipv6-flow-label-00.txt

Status of this memo

Abstract

The IPv6 flow label field has been designed to allow eliminating protocol layer violations and the related problems from flow specific packet classifiers.

This document provides an analysis of the current state of the IPv6 flow label field definition and proposes new text to be included in the next revision of the IPv6 specification.

Table of Contents

1. Introduction

1.1 Overview

At the time when the IPv6 specification [RFC2460] was written, the requirements for flow label field usage were still evolving.

The last several years of work in IETF provide new perspective and framework for the standardization of the IPv6 flow label. Also, the new charter of the IPv6 Working Group invites contributions to flow label standardization.

A detailed problem statement is provided in section 1.2, and the goals for the flow label definition in 1.3. Section 2 provides an analysis of the current definition of the IPv6 flow label [RFC2460, RFC1809, RFC2205]. Section 3 details the new definition with its implications, with proposed text to be included in the next revision of the IPv6 specification. Finally, section 4 provides some useful background information on the topic.

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.

1.2 Problem Statement

The IPv6 flow label field is designed to enable efficient classification of packets that should receive some flow-specific "special handling".

Without the flow label, flow classification must be based on the transport header information (port numbers). Snooping in the transport header is problematic due to several factors: The transport header may be unavailable because of either fragmentation, or IPsec encryption. Usage of IPv6 extension headers will also make finding the transport header more expensive, even if it is available. Finally, reliance on the transport header information is a layer violation and hinders introduction of new transport layer protocols (e.g. SCTP).

Current non-normative text in the Appendix A of [RFC2460] seems to be specific to the Integrated Services service model, unnecessarily restricting future work on defining new state establishment methods, but at the same time falls short in enabling flow label based classification of RSVP defined end-to-end flows in all cases.

The current normative specification of the flow label field in [RFC2460] is providing inadequate guidance for different flow state establishment methods to be defined.

See section 2.2 for more detailed analysis of the problems with Appendix A definition.

1.3 Requirements

The IPv6 protocol specification SHOULD only state generic rules, if any, governing the use of the flow label field by any flow state establishment method, and MUST enable co-existence of different flow state establishment methods in IPv6 hosts and routers.

The space of possible flow state establishment methods SHOULD NOT be restricted to end-to-end signaling protocols. For example, the IPv6 protocol specification should allow for future definition of administratively provisioned flows (automated through e.g. COPS, or manual configuration).

The text in the IPv6 protocol specification SHOULD leave the specifics arising from different flow state establishment methods, and different models of using the flow label to the documents that specify those methods and models.

The models for the use of the flow label and their specific state establishment methods should enable eliminating the layer violations in flow specific packet classifiers, thus facilitating evolution of the higher protocol layers independent of the specific flow state establishment method.

The semantics-free nature of the flow label, when out of context of the source and destination addresses, SHOULD be maintained.

Changes to the current specification SHOULD be kept minimal, and backwards compatibility SHOULD be maintained.

1.4 Terminology

Classifier	An entity which selects packets based on the content of packet headers according to defined rules.
Control plane	Part of the router taking care of router control functions, such as routing protocols and flow set-up signaling protocols. Controls the functions of the forwarding plane.
Forwarding plane	Part of the router receiving and forwarding user packets; also known as "fast path".

Multi-Field (MF) Classifier	A classifier which selects packets based on the content of some arbitrary number of header fields.

2. Analysis of the RFC 2460 Flow Label Specification

2.1 RFC 2460 Definition of the Flow Label

The IPv6 Flow Label is defined in [RFC2460] as a 20 bit field in the IPv6 header which may be used by a source to label sequences of packets for which it requests special handling by the IPv6 routers, such as non-default quality of service or "real-time" service.

The Flow Label aspect of IPv6 is stated to be "still experimental and subject to change". The only rule set for the flow label in the main body of the [RFC2460] is that if the Flow Label field use is not supported, it is set to zero when originating the packet, passed on unchanged when forwarding the packet, and ignored when receiving the packet.

2.1.1 Appendix A - Semantics and Usage of the Flow Label Field

The characteristics of IPv6 flows and flow labels, and the rules that govern the flow label functions are further defined in [RFC2460] Appendix A (non-normative text).

Background information on the documented semantics can be found in [RFC1809].

According to [RFC2460], the nature of the special handling might be conveyed to the routers by a control protocol, such as a resource reservation protocol, or by information within the flow's packets themselves, e.g., in a hop-by-hop option.

For the purpose of this document the rules from the Appendix A of [RFC2460] are rearranged as follows:

(a) A flow is uniquely identified by the combination of a source address and a non-zero flow label.

(b) Packets that do not belong to a flow carry a flow label of zero.

(c) A flow label is assigned to a flow by the flow's source node.

(d) New flow labels must be chosen (pseudo-)randomly and uniformly from the range 1 to FFFFF hex. The purpose of the random allocation is to make any set of bits within the Flow Label field suitable for use as a hash key by routers, for looking up the state associated with the flow.

(e) All packets belonging to the same flow must be sent with the same

source address, destination address, and flow label.

(f) If packets of a flow include a Hop-by-Hop Options header, then they all must be originated with the same Hop-by-Hop Options

header contents (excluding the Next Header field of the Hop-by-Hop Options header).

(g) If packets of a flow include a Routing header, then they all must be originated with the same contents in all extension headers up to and including the Routing header (excluding the Next Header field in the Routing header).

(h) The routers or destinations are permitted, but not required, to verify that these conditions are satisfied. If a violation is detected, it should be reported to the source by an ICMP Parameter Problem message, Code 0, pointing to the high-order octet of the Flow Label field (i.e., offset 1 within the IPv6 packet).

(i) The maximum lifetime of any flow-handling state established along a flow's path must be specified as part of the description of the state-establishment mechanism, e.g., the resource reservation protocol or the flow-setup hop-by-hop option.

(j) A source must not reuse a flow label for a new flow within the maximum lifetime of any flow-handling state that might have been established for the prior use of that flow label. When a node stops and restarts (e.g., as a result of a "crash"), it must be careful not to use a flow label that it might have used for an earlier flow whose lifetime may not have expired yet.

2.2 Applicability of the RFC 2460 Flow Label Definition

As stated in [RFC2460], the motivation for the Flow Label field is to request special handling for the packets belonging to a flow. The flow label value itself has no semantics, but it can be used by routers to determine the appertaining of a packet to a certain flow (classification), and to find the state containing the definition for the "special handling" admitted for that flow. The exact nature of the "special handling" is defined through means other than the flow label itself.

At the time of the definition of the rules in [RFC2460] the major method for defining the "special handling" by routers was the resource reservation signaling protocol (RSVP) of the "Integrated Services" architecture [RFC1633, RFC2205]. With the hindsight it seems that some of the flow label rules set in [RFC2460] Appendix A are quite specific to the Integrated Services model, and block the way forward for definition of other flow state establishment methods.

Some specific points are raised in the following subsections.

2.2.1 Lacking Support for RSVP WF Reservation Style

The Wildcard-Filter (WF) reservation style allows the RSVP session destination to reserve resources for transmission by any of the senders of the RSVP session. All the state that can be utilized for packet classification with the WF-style session is in the RSVP Session object, since the WF-style reservations have no Filter Specs [RFC2205].

Currently, for end-to-end flows, the Session object can only be specified in the terms of the destination address, transport protocol identifier (Id), and the destination port number. This results in layer violation in packet classification with all the identified problems (inefficiency, fragmentation, IPsec) in spite of using the flow label Filter Specs.

For WF-style sessions this situation could be remedied with a new type (or "C-Type") of a Session object, where only the destination IPv6 address and the flow label are specified (quite much like the Session object defined in [RFC3175] for aggregated flows). For other flow styles it might be more appropriate to have the session object to specify the destination address only, and have each Filter Spec to contain the source's flow label.

The problem with the current flow label rules is that if the flow label is set according to information received from the destination (e.g. through the Session Announcement Protocol (SAP)), it becomes possible that the same flow label value should be used for two different flows from the same source simultaneously. This can happen if the source is taking part to two different sessions with different destinations, and the flow label number generators in the destinations happen to pick up the same number. This is in direct contrast with the rule (a) above.

Requesting the destination to pick another flow label would be infeasible, if the RSVP sessions already have other senders.

2.2.2 Too Restricted Flow Classifier

The rule (a) states that a flow is uniquely identified by the source address and the flow label. Rule (e) states that all the packets in the flow must also have the same destination address.

This means that the source may not use the same flow label value for flows to two different destination addresses. As stated above, this is in violation with RSVP model, but also is unnecessarily hindering introduction of other flow state establishment methods.

2.2.3 RSVP/Integrated Services Specific Rules

The rules (f) and (g) state that any hop-by-hop or routing headers must be the same for all packets in a flow. This matches the Intserv practice of nailing down the path for the flow. The intent has probably been to enable the router to by-pass next-hop look-up and to supply pre-processed routing header contents for the packets in the flow.

These rules are clearly specific to the RSVP flow state establishment method, and should not be required of flows in general. The RSVP flow state establishment method would still mandate these rules for Integrated Services flows. The forwarding path of a IntServ capable router would also be able to enforce these rules for IntServ flows.

2.2.4 Too Restricting Rule for Flow Label Value Re-use

The rule (j) defines guard periods on re-use of flow label values. This is too restrictive even in the case of Intserv flows. There would be no harm in a (rebooted) node reusing a flow label value, as the RSVP signaling would enable the routers to flush any old state for the same flow classifier.

2.2.5 Unnecessary Rule for Flow Label Value Selection

Finally, it has been concluded that routers can't actually rely on the random distribution of the flow label values as required by the rule (d). In practice routers MUST be able to utilize algorithms that do not depend on the statistical distribution of the flow label values. Therefore, the rule (d) SHOULD be relaxed for flow labels in general. However, specific flow state establishment methods MAY still use pseudo-random numbers as flow label values.

2.2.6 Ambiguity on the End-to-End Nature of the Flow Label

The RSVP/Intserv usage calls for end-to-end immutable flow classifier. At the same time, the flow label field has been left unprotected by the Authentication Header (AH) computation.

3. New Flow Label Specification

The section proposes new text to be included in the IPv6 Specification. The section 3.1 is intended to provide a new version of the text in [RFC2460] section 6. The text in sections 3.2 and 3.3 could go to different parts of the IPv6 specification.

3.1 Proposed Flow Label Text for IPv6 Specification

The 20-bit Flow Label field in the IPv6 header MAY be used by a source to label sequences of packets for which it requests special handling. A non-zero flow label indicates that the IPv6 packet is labeled. IPv6 nodes receiving a labeled IPv6 packet can use the Source Address, Flow Label, Destination Address triplet to classify the packet to a certain flow. A flow is given some specific treatment based on the flow state established on a set of IPv6 routers. The nature of the specific treatment and the methods for the flow state establishment are out of scope for this specification.

The host MUST keep track of the Flow Label values in use to avoid trying to establish conflicting flow state. The Flow Label value, when set, is end-to-end immutable, but MAY be temporarily changed, if so required by the flow state establishment method.

Hosts or routers that do not support the functions of the Flow Label field MUST 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.

3.2 Requirements for Flow State Establishment Methods

The following MUST be considered by all flow state establishment methods:

(1) A flow is uniquely identified by the combination of the source address, non-zero flow label, and the destination address.

(2) All flow state MUST be created with a flow state establishment method. All such methods are out of scope for this specification.

(3) Flow state with the flow label value zero SHALL NOT be created.

(4) Host implementations SHOULD keep track of the flow label values used by any flow establishment methods from a local source address to all used destination addresses.

(5) A non-zero flow label value is end-to-end immutable, but can be changed if the established flow state so requires.

(6) The maximum lifetime of any established flow state must be specified as part of the flow state establishment method.

(7) A source MAY reuse a flow label value any time, unless otherwise specified by the flow state establishment method for the new flow.

(8) All flow state establishment methods MUST allow for the case where a router determines the offered flow to be in conflict with flow state created with an other flow state establishment method. If a conflict is detected, it SHOULD be reported by the flow state establishment method.

3.3 Implications of the New Definition

- The same flow label value can be used for different flows with the same source addresses, provided that the destination addresses are different (1).

- Each (source address, flow label, destination address) triplet can uniquely determine a flow and the relevant flow state, if any (1). However, multiple different triplets MAY determine the same flow, and refer to the same flow state, depending on how the flow was defined with the flow state establishment method.

- The only requirement for a flow label value used for the flow is that it MUST be non-zero (3). However, specific flow state establishment methods MAY use (non-zero) pseudo-random numbers as flow label values.

- A non-zero flow label value is guaranteed to be received by the destination. If the source sends the packet with a zero flow label value, a router in the network MAY set the flow label value to a non-zero value (5).

- A non-zero flow label value MAY be changed in transit by a router, but the original value MUST be restored before the packet leaves the domain of the flow state establishment method defining such a temporary change (5).

- Flow state lifetime may also be indefinite, if so specified by a flow state establishment method. The method MUST also provide the means to guarantee no dangling state (6).

- If new flow state is signaled through a certain path, the routers can flush any old state they might have, and install the new flow state (7).

- Some host implementations of flow state establishment methods might be impractical or impossible to synchronize in a host environment (e.g., the host OS may implement one method in the kernel with no interface to user space, where another state establishment method

is residing). Therefore the routers MUST be able to return an error as part of the flow establishment response, if the offered flow is deemed conflicting with a flow state created by another flow establishment method. Local policy at the router MAY set a precedence between the flow establishment methods, and MAY be able to cancel a lower precedence flow in favor of the new flow (8).

4. Conceptual Models Relating to the Flow Label

4.1 About Packet Classification

This section briefly summarizes issues relating to packet classification relating to the use of the flow label.

Packet classification happens in a context of an agreement ("contract") between a "customer" and a "provider". The only fields in the IP packet header that the provider can utilize in mapping a packet to a specific customer's contract are the source and destination address fields. We call this "customer classification". Other information, such as incoming link, can also be used to map packets to the customer's contract.

In the context of the contract governing the packet, the packet can be further classified to a flow. The packet filter rules for the flow classifiers in the network are part of the flow state. The flow state also specifies what kind of "special handling" the packets of the flow should get, and what are the flow traffic parameters (e.g. bandwidth, delay, etc.) and contains the flow usage counters.

It should be noted that the actual values of the header fields specified in a flow classifier are immaterial to the network operator - the operator assigns no specific semantics to any of the fields.

Actual implementations will likely combine the "customer classification" and flow classification into one filter rule, but the conceptual separation between the two is essential.

Usage of the IPv6 flow label greatly simplifies the filter rules, as the classification can be done on the basis of the IP addresses and the flow label alone.

A packet classified to a flow can be further mapped to a Behavior Aggregate (BA), enabling other routers in the network bypass flow classification. The Differentiated Services Code Point (DSCP) field is used to identify the selected BA [RFC2475].

4.2 Host Considerations for the Flow Label

4.2.1 Choosing Flow Label Values

A specific flow state establishment method MAY set requirements on the flow label values to be used. In any case, the host

implementation SHOULD keep track of the actual flow label values being used between a local source address and any destination addresses. The same facility keeping track of the flow label values SHOULD be utilized to check whether the flow label value chosen by

the flow state establishment method is currently in use or not between the given source and destination address pair, and SHOULD also return a flow label value assigned by host implementation specific algorithm, if the flow state establishment method did not specify any specific value.

4.2.2 End-to-End Negotiation

Flow label values for flows SHOULD be included as part of any end-to-end signaling dealing with the flow, e.g. RSVP for resource reservation, or SIP/SDP for end-to-end session establishment.

RSVP usage is analogous to the familiar MF classifier, but now the flow label replaces the need to specify the transport protocol and port numbers for the flow classifier.

In the case of SIP either the source or the destination could have a preference for the flow label value to be used. For example, the destination could have an agreement with its access provider effecting flow state for "special handling" for all packets marked with a certain flow label value towards the destination. Therefore the source SHOULD honor the destination's request to mark the packets with the flow label value specified.

4.2.3 Relation to the Other Packet Header Fields

A flow can be uniquely identified by the (source address, flow label, destination address) triplet. Any possible constraints for the rest of the IPv6 header fields or extension headers are to be specified by the flow state establishment method defining the flow semantics.

Flow state establishment methods SHOULD include the Mobile IP Home Addresses of the source and the destination in the state establishment process, if available. This enables avoiding state duplication on fixed portions of the path when either end changes its Care-of Address.

4.3 Router Considerations for the Flow Label

4.3.1 Flow Label is End-to-End Immutable

Routers MAY NOT change the end-to-end flow label value, unless explicitly so requested by the flow state establishment method. The flow state establishment method MUST be able to tell the destination which value to expect on the received packets.

Also, an administrative domain MAY internally change the flow label value, but it SHALL restore the original value on domain egress. Intra-domain modification MUST NOT interfere with inter-domain flow

set-up signaling carrying the original end-to-end immutable flow label value.

4.3.2 Flow Label Values Have No Known Properties

The router MUST NOT assume any specific property on the flow label values assigned by hosts. Router performance SHOULD NOT be dependent on the overall distribution of the flow label values of the established flows.

4.3.3 Conceptual Model for Flow State

This section lays out a simple conceptual model for minimal flow state in the router forwarding plane. Actual implementations may choose any implementation methods they like.

Router forwarding plane needs to maintain at least the following information (flow state) for each defined flow:

Source Address, Flow Label, Destination Address	The triplet identifying the flow.
Flow Accounting Information	Counter of the number of bytes or packets of the flow data forwarded. The router control plane can see from this if the flow has been active (since it was last checked), and how much data has been forwarded (useful for accounting purposes).
Forwarding Treatment	Defines the actual "special handling" the flow packets are subjected to.

The flow state is created by the router control plane via a flow state establishment method. The flow state establishment method definitions are out of scope for this specification.

Stale flow state is deleted by the router control plane after the flow expires, or when a new flow state overriding the old is created. The flow state can also be explicitly deleted via the flow state establishment method.

4.3.4 Classification

Packet classification is done by the router forwarding plane on the

flat 20-bit flow label, and the source and destination address fields.

When matching flow state has been found, the router will be able to update the Flow Accounting Information and forward the packet with the "special handling" as specified by the Forwarding Treatment in the flow state.

If flow state can not be located for a packet it is forwarded as if the flow label was zero, but the flow label is left intact. No flow state is maintained for unknown flows.

Appendix A: Why no Flow Label Format?

The choice to not introduce any internal format for the flow label represents the "minimal modification" policy and is intended to ease the process of the acceptance of this specification by the IPv6 community.

This is also in line with the removal of any "flags" from the IPv6 main header design, and the recent deprecation of the term "format prefix" in conjunction of IPv6 addresses.

In the same way as next-hop lookup should function independent of any "format prefixes" or administrative boundaries in the IPv6 addresses, the flow label lookup should remain independent of any possible internal structure for the flow label values themselves.

Abstaining from eating in to the 20 bits of the flow label also keeps maximal possibilities open for future refinement of this specification.

Appendix B: Why no Pseudo-Random Values?

[RFC2460] motivates the requirement for pseudo random flow label field with easing the hash key computation in routers doing flow classification. Hashing has to deal with the problem of large hash buckets due to unbalanced hash key distribution. If the router trusts on the hosts to generate good hash keys, it places itself on the mercy of the hosts. A not-so-good generator in any widely used host platform may become problematic for the router.

In recent years the hardware implementations of the classifiers have advanced, and schemes like search trees and Content Addressable Memory (CAM) are widely used for classification. It is the authors' view that the flow label specification SHOULD NOT favor any individual classification implementation strategy, especially when it provides no functional value for the purpose of the flow label itself, and seems to hinder future use of the flow label for non-signaled flow state establishment methods.

Hash implementations in routers can compute a hash key over the (source address, flow label, destination address) triplet.

References

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

[RFC1809] C. Partridge, "Using the Flow Label Field in IPv6", RFC 1809, June 1995.

[RFC2475] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Service", RFC 2475, December 1998.

[RFC1633] R. Braden, D. Clark, S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633, June 1994.

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

[RFC3175] F. Baker, C. Iturralde, F. Le Faucheur, B. Davie, "Aggregation of RSVP for IPv4 and IPv6 Reservations", RFC 3175, September 2001.

[Conta] A. Conta, B. Carpenter, "A proposal for the IPv6 Flow Label Specification", Internet Draft <draft-conta-ipv6-flow-label-02.txt>, July 2001, expires January 2002, Work in progress.

Security Considerations

Anything that facilitates flow classification also increases the vulnerability to traffic analysis.

The use of flow label in general enables flow classification also in the presence of ESP headers. This allows the transport header values to remain confidential, which may lessen the possibilities for some forms of traffic analysis.

Acknowledgements

The discussion on the topic in the IPv6 WG mailing list has been instrumental for the definition of this specification. The authors want to thank Steve Blake, Jim Bound, Brian Carpenter, Francis Dupont, Robert Elz, Tony Hain, Christian Huitema, Frank Kastenholz, Hesham Soliman, Michael Thomas for their tireless contributions on the list.

Charles Perkins reviewed the text and provided many helpful comments.

Author's Address

Jarno Rajahalme
Nokia Research Center
P.O. Box 407
FIN-00045 NOKIA GROUP,
Finland
E-mail: jarno.rajahalme@nokia.com

Alex Conta
Transwitch Corporation
3 Enterprise Drive
Shelton, CT 06484
USA
Email: aconta@txc.com

Expiration Date

This memo is filed as <draft-rajahalme-ipv6-flow-label-00.txt> and expires in May 2002.