rtgwg D. Lamparter
Internet-Draft NetDEF
Intended status: Standards Track October 20, 2014
Expires: April 23, 2015

# Considerations and Registry for extending IP route lookup
draft-lamparter-rtgwg-routing-extra-qualifiers-00

## Abstract

This document describes the behaviour of a routing system that takes additional specifications on routes--extra qualifiers--into account on a hop-by-hop basis, augmenting longest match behaviour.

## Status of This Memo



## Copyright Notice



## Table of Contents

# 1. Introduction

IP Routing systems at the time of creation of this document are occasionally already capable of matching more than the packet’s destination addresses to lookup routes, preexisting patterns include virtual routers (i.e. keying by routing instance), QoS-aware routing (keying by DSCP bits) and the relatively unspecific "policy routing."

Additional developments extend this field to the point where a lack of well-defined specification may lead to interoperability problems. The intent of this document is to construct a reference framework for extensions on the match aspect of IP routes.

Specifically, since IP Routing includes longest-match route selection, the ordering of all match qualifiers must be the same among all routers to prevent loops or connectivity loss.

While this document is written with IPv6 in mind, it applies to IP router architecture in general, including IPv4 routers.

## 1.1. Requirements Language

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. Applicability

While the conceptually same longest-prefix routing is used not only for routing packets, but also recursive route/nexthop lookups, multicast reverse path forwarding and unicast reverse path filtering. However, while based on the same base principle, these applications may differ in their requirements. For example, multicast RPF cannot use source address discriminators since no source address is known at the time of lookup.

The intent of this specification is only to provide a basic framework; individual extensions to route match behaviour MUST clarify their respective applicability.

## 3. Match criteria (informational)

### 3.1. Virtual routers

While not documented to this extent, an implementation capable of partitioning a physical router into multiple virtual routers is an application that essentially has the virtual router identifier as first key in lookup operations. This may not be implemented as such, for example by keeping tables completely separate, however the end behaviour is the same; lookups are made local to the router instance.

### 3.2. Policy routing

Equally little specified as virtual routers, policy routing usually applies certain qualifiers (source address, traffic class, firewall markers) prior to destination address match.

### 3.3. Destination address longest match

The conventional destination IP address longest match is included at this point as it is, barring implementation specific extensions mentioned above, the first qualifier used to match packets against the route table.

### 3.4. Source address longest match

Currently under development, matching on the source address permits routers to choose the correct (in terms of [RFC2827]) exit in smaller multihomed networks.  This is distinct from policy routing in that only few select (usually default) routes would be annotated with source prefixes.

Various aspects of this are described in:

[I-D.troan-homenet-sadr]

[I-D.boutier-homenet-source-specific-routing]

[I-D.sarikaya-6man-sadr-overview]

[I-D.baker-rtgwg-src-dst-routing-use-cases]

[I-D.baker-ipv6-isis-dst-src-routing]

[I-D.baker-ipv6-ospf-dst-src-routing]

[I-D.baker-rtgwg-src-dst-routing-use-cases]

### 3.5.  Flowlabel routing

TBD, described in:

[I-D.baker-ipv6-isis-dst-flowlabel-routing]

[I-D.baker-ipv6-ospf-dst-flowlabel-routing]

### 3.6.  QoS/DSCP traffic class based routing

TBD (deprecated, reference only)

## 4.  Requirements to extending match behaviour

### 4.1.  Match ordering

Adding further criteria to be looked up when forwarding packets on a hop-by-hop basis has the very fundamental requirement that all routers behave the same way in choosing the most specific route when there are multiple eligible routes.

This document disambiguates this situation by recording the order of specificness in a registry.  This means that the comparison for "more specific", here indicated by A < B (to mean A is more specific than B), is redefined as concatenation for attributes a, b, c as:

```
A < B :=    Aa <  Ba
        || (Aa == Ba && Ab <  Bb)
        || (Aa == Ba && Ab == Bb && Ac < Bc )
```

This transfers to a sample situation (using source address, destination address and flowlabel as qualifiers):

Example route table

| | destination | source | flowlabel |
|---|---|---|---|
| route A: | 2001:db8::/32 | | |
| route B: | 2001:db8:1234::/48 | 2001:db8:4567::/48 | |
| route C: | 2001:db8:1234::/48 | | abcde |
| route D: | 2001:db8:1234:5678::/64 | 2001:db8:4567::/48 | abcde |
| route E: | 2001:db8:1234:5678::/64 | | |

Showing the different results between "destination, source, flowlabel" ("DSF") and "destination, flowlabel, source" ("DFS") ordering:

Example match results

| packet to be routed | | | | result | |
|---|---|---|---|---|---|
| # | destination | source | flowlabel | "DSF" | "DFS" |
| 1 | 2001:db8::1 | 2001:db8:4567::1 | abcde | A | A |
| 2 | 2001:db8:1234::1 | 2001:db8:4567::1 | abcde | B | C |
| 3 | 2001:db8:1234::1 | 2001:db8:4567::1 | 11111 | B | B |
| 4 | 2001:db8:1234::1 | 2001:db8:1111::1 | abcde | C | C |
| 5 | 2001:db8:1234::1 | 2001:db8:1111::1 | 11111 | A | A |
| 6 | 2001:db8:1234:5678::1 | 2001:db8:4567::1 | abcde | D | D |
| 7 | 2001:db8:1234:5678::1 | 2001:db8:4567::1 | 11111 | E | E |
| 8 | 2001:db8:1234:5678::1 | 2001:db8:1111::1 | abcde | E | E |

It should be noted that lookup may not result in usage of the most specific element even for the first attribute (destination in the example). As displayed in #5 above, the route used is the most specific one that satisfies all conditions. If a system cannot "back out" to less specific matches on earlier attributes, this MUST be worked around by installing synthetic routes for these cases.

## 4.2. Compatibility / Interoperability

Since a router implementing extra match qualifiers can have additional, more specific routes than one that doesn't implement these qualifiers, persistent loops can form between these systems. To prevent this from happening, a simple rule must be followed:

The set of qualifiers used to route a particular packet MUST be a subset of the qualifiers supported by the next hop.

This means in particular that a router using extra qualifier A MUST NOT route packets based on a route that checks this qualifier to a system that doesn't support qualifier A (and hence doesn't understand the route).

There are 3 possible approaches to avoid such a condition:

1. discard the packet (treat as destination unreachable)

2. calculate an alternate topology including only routers that support qualifier A

3. if the lookup returns the same nexthop without using qualifier A, use that result (i.e., the nexthop is known to correctly route the packet)

Above considerations require under all circumstances a knowledge of the next router's capabilities. For routing protocols based on hop-by-hop flooding (RIP [RFC2080], BGP [RFC4271]), knowing the peer's capabilities - or simply relying on systems to only flood what they understand - is sufficient. Protocols building a link-state database (OSPF [RFC5340], IS-IS [RFC5308]) have the additional opportunity to calculate alternate paths based on knowledge of the entire domain, but cannot rely on routers flooding only link state they support themselves.

## 5. IANA Considerations

This document requests creation of a new registry called the "Routing Qualifier Registry." The registry consists of an ordered list of items, no identifier value needs to be assigned. The only purpose of the registry is to document the order in which qualifiers are evaluated.

Registry items must specify the following information:

- Name of the qualifier

- Applicable protocols (IP version 4 and/or IP version 6)

- o Specification reference (possibly distinct between IPv4 and IPv6)

- o Insertion position, listing both the previous and next entry to avoid confusion

The allocation policy per [RFC5226] is "IETF Review." This is intended to help keep routing systems compatible with each other.

### 5.1. Initial list

The list is prepropagated with a single entry describing "classical" destination-based routing:

Name: Destination lookup

Applicable to IPv4 and IPv6

Specification references: [RFC4632] for IPv4, [RFC2460] for IPv6

## 6. Security Considerations

This document specifies only the ordering of lookups. Making no change to the existing situation, there are no security considerations for this document.

## 7. Privacy Considerations

As with security considerations, no privacy considerations apply to this document.

Introducing additional routing qualifiers has the potential to expose information that was not previously visible, in particular if such information would otherwise be scrubbed by a process like NAT. However, these considerations are left for documents actually introducing new routing qualifiers.

## 8. Acknowledgements

This document is largely the result of discussions with Fred Baker.

A lot of drafts exists in this general area, refer to the informative references section below.

# 9. Change Log

Initial Version: October 2014

# 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.

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

[RFC4632] Fuller, V. and T. Li, "Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan", BCP 122, RFC 4632, August 2006.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

## 10.2. Informative References

[I-D.baker-ipv6-isis-dst-flowlabel-routing]
Baker, F., "Using IS-IS with Token-based Access Control", draft-baker-ipv6-isis-dst-flowlabel-routing-01 (work in progress), August 2013.

[I-D.baker-ipv6-isis-dst-src-routing]
Baker, F., "IPv6 Source/Destination Routing using IS-IS", draft-baker-ipv6-isis-dst-src-routing-01 (work in progress), August 2013.

[I-D.baker-ipv6-ospf-dst-flowlabel-routing]
Baker, F., "Using OSPFv3 with Token-based Access Control", draft-baker-ipv6-ospf-dst-flowlabel-routing-03 (work in progress), August 2013.

[I-D.baker-ipv6-ospf-dst-src-routing]
Baker, F., "IPv6 Source/Destination Routing using OSPFv3", draft-baker-ipv6-ospf-dst-src-routing-03 (work in progress), August 2013.

[I-D.baker-rtgwg-src-dst-routing-use-cases]
Baker, F., "Requirements and Use Cases for Source/Destination Routing", draft-baker-rtgwg-src-dst-routing-use-cases-00 (work in progress), August 2013.

[I-D.boutier-homenet-source-specific-routing]
Boutier, M. and J. Chroboczek, "Source-specific Routing", draft-boutier-homenet-source-specific-routing-00 (work in progress), July 2013.

[I-D.sarikaya-6man-sadr-overview]
Sarikaya, B., "Overview of Source Address Dependent Routing", draft-sarikaya-6man-sadr-overview-01 (work in progress), September 2014.

[I-D.troan-homenet-sadr]
Troan, O. and L. Colitti, "IPv6 Multihoming with Source Address Dependent Routing (SADR)", draft-troan-homenet-sadr-01 (work in progress), September 2013.

[RFC2080] Malkin, G. and R. Minnear, "RIPng for IPv6", RFC 2080, January 1997.

[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.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, October 2008.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.

## Author's Address

David Lamparter
NetDEF
Leipzig  04103
Germany

Email: david@opensourcerouting.org