Internet Engineering Task Force (IETF)
Request for Comments: 8028
Updates: 4861
Category: Standards Track
ISSN: 2070-1721

F. Baker
B. Carpenter
Univ. of Auckland

November 2016

First-Hop Router Selection by Hosts in a Multi-Prefix Network

Abstract

This document describes expected IPv6 host behavior in a scenario that has more than one prefix, each allocated by an upstream network that is assumed to implement BCP 38 ingress filtering, when the host has multiple routers to choose from. It also applies to other scenarios such as the usage of stateful firewalls that effectively act as address-based filters. Host behavior in choosing a first-hop router may interact with source address selection in a given implementation. However, the selection of the source address for a packet is done before the first-hop router for that packet is chosen. Given that the network or host is, or appears to be, multihomed with multiple provider-allocated addresses, that the host has elected to use a source address in a given prefix, and that some but not all neighboring routers are advertising that prefix in their Router Advertisement Prefix Information Options, this document specifies to which router a host should present its transmission. It updates RFC 4861.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc8028.

Table of Contents

1. Introduction and Applicability

This document describes the expected behavior of an IPv6 [RFC2460] host in a network that has more than one prefix, each allocated by an upstream network that is assumed to implement BCP 38 [RFC2827] ingress filtering, and in which the host is presented with a choice of routers. It expects that the network will implement some form of egress routing, so that packets sent to a host outside the local network from a given ISP's prefix will go to that ISP. If the packet is sent to the wrong egress, it is liable to be discarded by the BCP 38 filter. However, the mechanics of egress routing once the packet leaves the host are out of scope. The question here is how the host interacts with that network.

Various aspects of this issue, and possible solution approaches, are discussed in "IPv6 Multihoming without Network Address Translation" [RFC7157].

BCP 38 filtering by ISPs is not the only scenario where such behavior is valuable. Implementations that combine existing recommendations, such as [RFC6092] and [RFC7084] can also result in such filtering. Another case is when the connections to the upstream networks include stateful firewalls, such that return packets in a stream will be discarded if they do not return via the firewall that created the state for the outgoing packets. A similar cause of such discards is unicast reverse path forwarding (uRPF) [RFC3704].

In this document, the term "filter" is used for simplicity to cover all such cases. In any case, one cannot assume that the host is aware whether an ingress filter, a stateful firewall, or any other type of filter is in place. Therefore, the only known consistent solution is to implement the features defined in this document.

Note that, apart from ensuring that a message with a given source address is given to a first-hop router that appears to know about the prefix in question, this specification is consistent with [RFC4861]. Nevertheless, implementers of Sections 6.2.3, 6.3.4, 6.3.6, and 8.1 of RFC 4861 should extend their implementations accordingly. This specification is fully consistent with [RFC6724] and depends on support for its Rule 5.5 (see Section 3.3). Hosts that do not support these features may fail to communicate in the presence of filters as described above.

1.1. Host Model

It could be argued that the proposal in this document, which is to send messages using a source address in a given prefix to the router that advertised the prefix in its Router Advertisement (RA), is a form of the Strong End System (ES, e.g., Host) model, discussed in Section 3.3.4.2 of [RFC1122]. In short, [RFC1122] identifies two basic models. First, the "strong host" model describes the host as a set of hosts in one chassis, each of which uses a single address on a single interface and always both sends and receives on that interface. Alternatively, the "weak host" model treats the host as one system with zero or more addresses on every interface and is capable of using any interface for any communication. As noted there, neither model is completely satisfactory. For example, a host with a link-local-only interface and a default route pointing to that interface will necessarily send packets using that interface but with a source address derived from some other interface, and will therefore be a de facto weak host. If the router upstream from such a host implements BCP 38 Ingress Filtering [RFC2827], such as by implementing uRPF on each interface, the router might prevent communication by weak hosts.

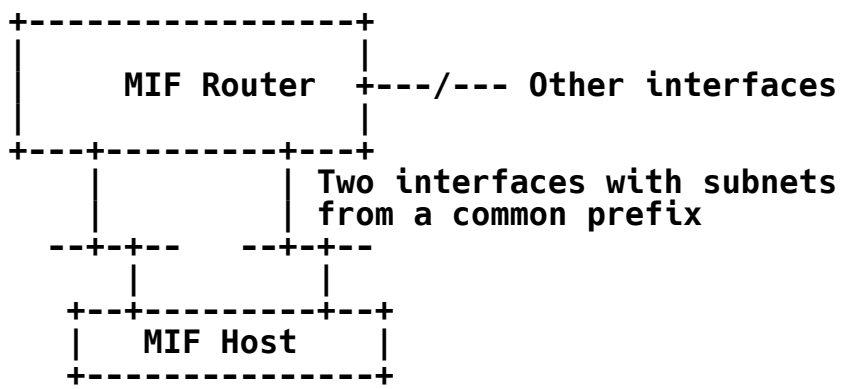

Figure 1: Hypothetical MIF Interconnection

The proposal also differs slightly from the language in [RFC1122] for the Strong Host model. The proposal is that the packet will go to a router that advertised a given prefix but that does not specify what interface that might happen on. Hence, if the router is a multi-interface (MIF) router and it is using a common prefix spanning two or more LANs shared by the host (as in Figure 1), the host might use either of those LANs, according to this proposal. The Strong Host model is not stated in those terms, but in terms of the interface used. A strong host would treat such an MIF router as two separate routers when obeying the rules from RFC 1122 as they apply in the Strong case:

(A) A host MUST silently discard an incoming datagram whose destination address does not correspond to the physical interface through which it is received.

(B) A host MUST restrict itself to sending (non-source-routed) IP datagrams only through the physical interface that corresponds to the IP source address of the datagrams.

However, when comparing the presumptive route lookup mechanisms in each model, this proposal is indeed most similar to the Strong Host model, as is any source/destination routing paradigm.

Strong: route (src IP addr, dest IP addr, TOS) -> gateway

Weak: route (dest IP addr, TOS) -> gateway, interface

In the hypothetical MIF model suggested in Figure 1, the address fails to identify a single interface, but it does identify a single gateway.

1.2. 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. Sending Context Expected by the Host

2.1. Expectations the Host Has of the Network

A host receives prefixes in a Router Advertisement [RFC4861], which goes on to identify whether they are usable by Stateless Address Autoconfiguration (SLAAC) [RFC4862] with any type of interface identifier [RFC4941] [RFC7217]. When no prefixes are usable for SLAAC, the Router Advertisement would normally signal the availability of DHCPv6 [RFC3315] and the host would use it to configure its addresses. In the latter case (or if both SLAAC and DHCPv6 are used on the same link for some reason), the configured addresses generally match one of the prefixes advertised in a Router Advertisement that are supposed to be on-link for that link.

The simplest multihomed network implementation in which a host makes choices among routers might be a LAN with one or more hosts on it and two or more routers, one for each upstream network, or a host that is served by disjoint networks on separate interfaces. In such a network, especially the latter, there is not necessarily a routing protocol, and the two routers may not even know that the other is a router as opposed to a host, or may be configured to ignore its

presence. One might expect that the routers may or may not receive each other's RAs and form an address in the other router's prefix (which is not per [RFC4862], but is implemented by some stub router implementations). However, all hosts in such a network might be expected to create an address in each prefix so advertised.

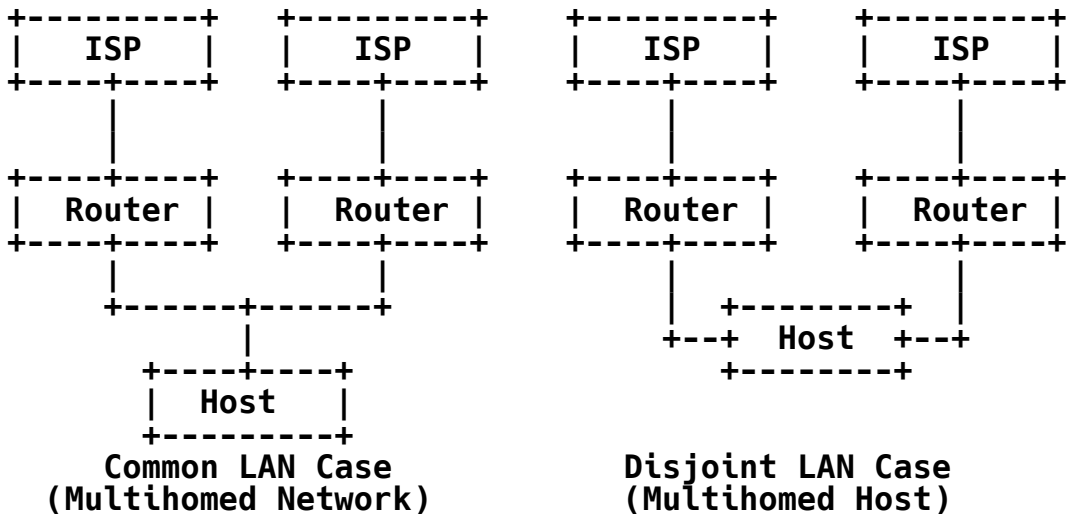

Figure 2: Two Simple Networks

If there is no routing protocol among those routers, there is no mechanism by which packets can be deterministically forwarded between the routers (as described in BCP 84 [RFC3704]) in order to avoid filters. Even if there was routing, it would result in an indirect route, rather than a direct route originating with the host; this is not "wrong", but can be inefficient. Therefore, the host would do well to select the appropriate router itself.

Since the host derives fundamental default routing information from the Router Advertisement, this implies that, in any network with hosts using multiple prefixes, each prefix SHOULD be advertised via a Prefix Information Option (PIO) [RFC4861] by one of the attached routers, even if addresses are being assigned using DHCPv6. A router that advertises a prefix indicates that it is able to appropriately route packets with source addresses within that prefix, regardless of the setting of the L and A flags in the PIO.

In some circumstances, both L and A might be zero. If SLAAC is not wanted (A=0) and there is no reason to announce an on-link prefix (L=0), a PIO SHOULD be sent to inform hosts that they should use the router in question as the first hop for packets with source addresses in the PIO prefix. An example case is the MIF router in Figure 1, which could send PIOs with A=L=0 for the common prefix. Although

this does not violate the existing standard [RFC4861], such a PIO has not previously been common, and it is possible that existing host implementations simply ignore such a PIO or that existing router implementations are not capable of sending such a PIO. Newer implementations that support this mechanism should be updated accordingly:

- A host SHOULD NOT ignore a PIO simply because both L and A flags are cleared (extending Section 6.3.4 of [RFC4861]).

- A router SHOULD be able to send such a PIO (extending Section 6.2.3 of [RFC4861]).

2.2. Expectations of Multihomed Networks

Networking equipment needs to support source/destination routing for at least some of the routes in the Forwarding Information Base (FIB), such as default egress routes differentiated by source prefix. Installation of source/destination routes in the FIB might be accomplished using static routes, Software-Defined Networking (SDN) technologies, or dynamic routing protocols.

3. Reasonable Expectations of the Host

3.1. Interpreting Router Advertisements

As described in [RFC4191] and [RFC4861], a Router Advertisement may contain zero or more Prefix Information Options (PIOs) or zero or more Route Information Options (RIOs). In their original intent, these indicate general information to a host: "the router whose address is found in the source address field of this packet is one of your default routers", "you might create an address in this prefix", or "this router would be a good place to send traffic directed to a given destination prefix". In a multi-prefix network with multiple exits, the host's characterization of each default router SHOULD include the prefixes it has announced (extending Section 6.3.4 of [RFC4861]). In other words, the PIO is reinterpreted to also imply that the advertising router would be a reasonable first hop for any packet using a source address in any advertised prefix, regardless of Default Router Preference.

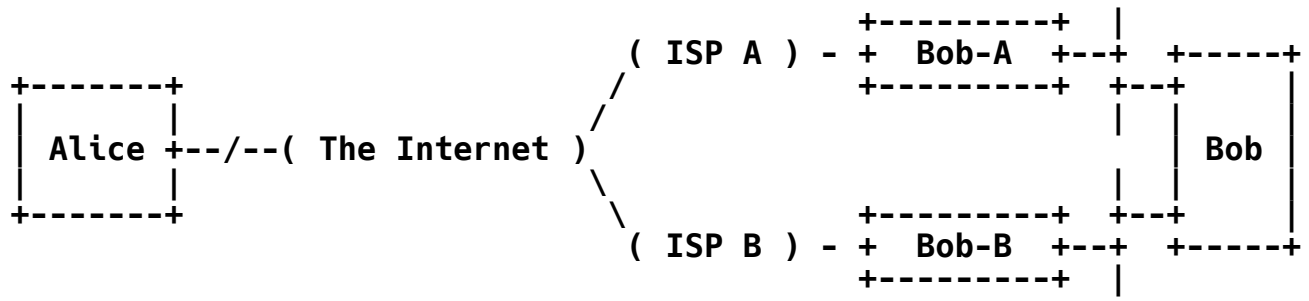

Figure 3: PIOs, RIOs, and Default Routes

The implications bear consideration. Imagine, Figure 3, that hosts Alice and Bob are in communication. Bob's network consists of at least Bob (the computer), two routers (Bob-A and Bob-B), and the links between them; it may be much larger, for example, a campus or corporate network. Bob's network is therefore multihomed, and Bob's first-hop routers are Bob-A (to the upstream ISP A advertising prefix PA) and Bob-B (to the upstream network B and advertising prefix PB). We assume that Bob is not applying Rule 5.5 of [RFC6724]. If Bob is responding to a message from Alice, his choice of source address is forced to be the address Alice used as a destination (which we may presume to have been in prefix PA). Hence, Bob either created or was assigned an address in PA, and can only reasonably send traffic using it to Bob-A as a first-hop router. If there are several routers in Bob's network advertising the prefix PA (referred to as "Bob-Ax" routers), then Bob should choose its first-hop router only from among those routers. From among the multiple Bob-Ax routers, Bob should choose the first-hop router based on the criteria specified in Section 3 of [RFC4191]. If none of the Bob-Ax routers has advertised an RA with a non-zero Router Lifetime or an RIO with a non-zero Route Lifetime that includes Alice, but router Bob-B has, it is irrelevant. Bob is using the address allocated in PA and courts a BCP 38 discard if he doesn't send the packet to Bob-A.

In the special case that Bob is initiating the conversation, an RIO might, however, influence source address choice. Bob could presumably use any address allocated to him, in this case, his address in PA or PB. If Bob-B has advertised an RIO for Alice's prefix and Bob-A has not, Bob MAY take that fact into account in address selection -- choosing an address that would allow him to make use of the RIO.

3.2. Default Router Selection

Default Router Selection (Section 6.3.6 of [RFC4861]) is extended as follows: A host SHOULD select default routers for each prefix it is assigned an address in. Routers that have advertised the prefix in their Router Advertisement message SHOULD be preferred over routers that do not advertise the prefix, regardless of Default Router Preference. Note that this document does not change the way in which default router preferences are communicated [RFC4191].

If no router has advertised the prefix in an RA, normal routing metrics will apply. An example is a host connected to the Internet via one router, and at the same time connected by a VPN to a private domain that is also connected to the global Internet.

As a result of this, when a host sends a packet using a source address in one of those prefixes and has no history directing it otherwise, it SHOULD send it to the indicated default router. In the "simplest" network described in Section 2.1, that would get it to the only router that is directly capable of getting it to the right ISP. This will also apply in more complex networks, even when more than one physical or virtual interface is involved.

In more complex cases, wherein routers advertise RAs for multiple prefixes whether or not they have direct or isolated upstream connectivity, the host is dependent on the routing system already. If the host gives the packet to a router advertising its source prefix, it should be able to depend on the router to do the right thing.

3.3. Source Address Selection

There is an interaction with Default Address Selection [RFC6724]. A host following the recommendation in the previous section will store information about which next hops advertised which prefixes. Rule 5.5 of RFC 6724 states that the source address used to send to a given destination address should, if possible, be chosen from a prefix known to be advertised by the next-hop router for that destination. Therefore, this selection rule SHOULD be implemented in a host following the recommendation in the previous section.

3.4. Redirects

There is potential for adverse interaction with any off-link Redirect (Redirect for a destination that is not on-link) message sent by a router in accordance with Section 8 of [RFC4861]. Hosts SHOULD apply off-link redirects only for the specific pair of source and destination addresses concerned, so the host's Destination Cache

might need to contain appropriate source-specific entries. This extends the validity check specified in Section 8.1 of [RFC4861].

3.5. History

Some modern hosts maintain history, in terms of what has previously worked or not worked for a given address or prefix and in some cases the effective window and Maximum Segment Size (MSS) values for TCP or other protocols. This might include a next-hop address for use when a packet is sent to the indicated address.

When such a host makes a successful exchange with a remote destination using a particular address pair, and the host has previously received a PIO that matches the source address, then the host SHOULD include the prefix in such history, whatever the setting of the L and A flags in the PIO. On subsequent attempts to communicate with that destination, if it has an address in that prefix at that time, a host MAY use an address in the remembered prefix for the session.

4. Residual Issues

Consider a network where routers on a link run a routing protocol and are configured with the same information. Thus, on each link, all routers advertise all prefixes on that link. The assumption that packets will be forwarded to the appropriate egress by the local routing system might cause at least one extra hop in the local network (from the host to the wrong router, and from there to another router on the same link).

In a slightly more complex situation such as the disjoint LAN case of Figure 2, for example, a home plus corporate home-office configuration, the two upstream routers might be on different LANs and therefore different subnets (e.g., the host is itself multihomed). In that case, there is no way for the "wrong" router to detect the existence of the "right" router, or to route to it.

In such a case, it is particularly important that hosts take the responsibility to memorize and select the best first hop as described in Section 3.

5. IANA Considerations

This document does not request any registry actions.

6. Security Considerations

This document is intended to avoid connectivity issues in the presence of BCP 38 ingress filters or stateful firewalls combined with multihoming. It does not, in itself, create any new security or privacy exposures. However, since the solution is designed to ensure that routing occurs correctly in situations where it previously failed, this might result in unexpected exposure of networks that were previously unreachable.

There might be a small privacy improvement: with the current practice, a multihomed host that sends packets with the wrong address to an upstream router or network discloses the prefix of one upstream to the other upstream network. This practice reduces the probability of that occurrence.

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <http://www.rfc-editor.org/info/rfc2460>.

[RFC4191] Draves, R. and D. Thaler, "Default Router Preferences and More-Specific Routes", RFC 4191, DOI 10.17487/RFC4191, November 2005, <http://www.rfc-editor.org/info/rfc4191>.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, DOI 10.17487/RFC4861, September 2007, <http://www.rfc-editor.org/info/rfc4861>.

[RFC6724] Thaler, D., Ed., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, DOI 10.17487/RFC6724, September 2012, <http://www.rfc-editor.org/info/rfc6724>.

7.2. Informative References

[RFC1122] Braden, R., Ed., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, DOI 10.17487/RFC1122, October 1989, <http://www.rfc-editor.org/info/rfc1122>.

[RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, DOI 10.17487/RFC2827, May 2000, <http://www.rfc-editor.org/info/rfc2827>.

[RFC3315] Droms, R., Ed., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, DOI 10.17487/RFC3315, July 2003, <http://www.rfc-editor.org/info/rfc3315>.

[RFC3704] Baker, F. and P. Savola, "Ingress Filtering for Multihomed Networks", BCP 84, RFC 3704, DOI 10.17487/RFC3704, March 2004, <http://www.rfc-editor.org/info/rfc3704>.

[RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, DOI 10.17487/RFC4862, September 2007, <http://www.rfc-editor.org/info/rfc4862>.

[RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, DOI 10.17487/RFC4941, September 2007, <http://www.rfc-editor.org/info/rfc4941>.

[RFC6092] Woodyatt, J., Ed., "Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service", RFC 6092, DOI 10.17487/RFC6092, January 2011, <http://www.rfc-editor.org/info/rfc6092>.

[RFC7084] Singh, H., Beebee, W., Donley, C., and B. Stark, "Basic Requirements for IPv6 Customer Edge Routers", RFC 7084, DOI 10.17487/RFC7084, November 2013, <http://www.rfc-editor.org/info/rfc7084>.

[RFC7157] Troan, O., Ed., Miles, D., Matsushima, S., Okimoto, T., and D. Wing, "IPv6 Multihoming without Network Address Translation", RFC 7157, DOI 10.17487/RFC7157, March 2014, <http://www.rfc-editor.org/info/rfc7157>.

[RFC7217] Gont, F., "A Method for Generating Semantically Opaque Interface Identifiers with IPv6 Stateless Address Autoconfiguration (SLAAC)", RFC 7217, DOI 10.17487/RFC7217, April 2014, <http://www.rfc-editor.org/info/rfc7217>.

Acknowledgements

Comments were received from Jinmei Tatuya and Ole Troan, who have suggested important text, plus Mikael Abrahamsson, Steven Barth, Carlos Bernardos Cano, Chris Bowers, Zhen Cao, Juliusz Chroboczek, Toerless Eckert, David Farmer, Bob Hinden, Ben Laurie, Dusan Mudric, Tadahisa Okimoto, Pierre Pfister, Behcet Sarikaya, Mark Smith, and James Woodyatt.

Authors' Addresses

Fred Baker
Santa Barbara, California 93117
United States of America

Email: FredBaker.IETF@gmail.com

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

Email: brian.e.carpenter@gmail.com