Internet Engineering Task Force
Internet-Draft
Updates: 4861 (if approved)
Intended status: Experimental
Expires: April 30, 2017

E. Kline
Google Japan KK
M. Abrahamsson
T-Systems Nordic
October 27, 2016

IPv6 Router Advertisement Prefix Information Option Exclusive Bit
draft-pioxfolks-6man-pio-exclusive-bit-01

Abstract

This document defines a new control bit in the IPv6 RA PIO flags octet that indicates that the node receiving this RA is the exclusive receiver of all traffic destined to any address within that prefix.

Termed the eXclusive bit (or "X bit"), nodes that recognize this can perform some optimizations to save time and traffic (e.g. disable ND and DAD for addresses within this prefix) and more immediately pursue the benefits of being provided multiple addresses (vis. [RFC7934] section 3). Additionally, network infrastructure nodes (routers, switches) can benefit by minimizing the number of {link layer, IP} address pairs required to offer network connectivity (vis. [RFC7934] section 9.3).

Use of the X bit is backward compatible with existing IPv6 standards compliant implementations.

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 April 30, 2017.

Table of Contents

1. Introduction

This document defines a new control bit in the Internet Protocol version 6 (IPv6) Router Advertisement (RA) Prefix Information Option (PIO) flags octet that indicates that the node receiving this RA is the exclusive receiver of all traffic destined to any address with that prefix. Subject to the lifetime constraints within the PIO, the receiving node effectively has exclusive use of the prefix, and will be the next hop destination for the sending router, and possibly other routers, for all traffic destined toward the prefix.

Termed the eXclusive bit (or "X bit"), nodes that recognize this can perform some optimizations to save time and traffic (e.g. disable Neighbor Discovery (ND) and Duplicate Address Detection (DAD) for addresses within this prefix) and more immediately pursue the benefits of being provided multiple addresses (vis. [RFC7934] section 3).

Additionally, network infrastructure nodes (routers, switches) can benefit by minimizing the number of {link layer, IP} address pairs required to offer network connectivity (vis. [RFC7934] section 9.3). A router, for example, need not create any {link layer, IP} address pair entries for IP address within a proffered exclusive-use prefix--it can reliably forward all traffic to the network node to which it advertised the prefix. This solves one potential link layer state exhaustion problem, i.e excessive number of {link layer, IP address pairs}, using IP layer forwarding.

Use of the X bit is backward compatible with existing IPv6 standards compliant implementations. [RFC4861]-compliant nodes that do not understand the X bit are not negatively impacted. They must ignore it, and can process the PIO under existing standards, making use of the information exactly as if the X bit were not set.

2. Motivation

This work is motivated by the pursuit of two categories of benefits: some host and network side improvements in efficiency, and support for new deployment architectures and address space use models.

2.1. Efficiency improvements

If a host knows it has exclusive use of a prefix it can perform some optimizations to save time an traffic. It can avoid ND on the receiving interface for addresses within these prefixes. Network interfaces can even drop Neighbor Solicitations for these addresses on the receiving interface to save power by not waking up more power hungry CPUs.

Additionally, a host can save time by not performing DAD for addresses within an exclusive-use prefix on the receiving interface. A host that wanted, for example, to use 2**64 unique IPv6 source addresses for DNS queries in order to improve resilience against forged answers (as recommended in section 9.2 of <RFC5452>), could do so without delaying each query from a newly formed address. A node could in theory implement the same strategy using Optimistic Duplicate Address Detection [1], but it could be very unfriendly to the network infrastructure (in terms of {link-layer, IP address} pair state) to do so without some explicit signal.

2.2. New architectural possibilities

There are several initiatives that propose network side practices that provide customer isolation, enhanced operational scalability, power efficiency, security and other benefits in IPv6 network deployments. Some of these involve isolating a host (or RA accepting client node) so that the host is the only node to receive a specific prefix, including

- DHCPv6 Prefix Delegation to hosts (<https://tools.ietf.org/html/draft-templin-v6ops-pdhost>), and

- advertising a unique prefix per host via unique RAs. (<https://tools.ietf.org/html/draft-ietf-v6ops-unique-ipv6-prefix-per-host>).

Some architectures further isolate the host layers below IPv6, for improved client node security.

Regardless of the specific level of isolation, the host can best make choices about its use of a prefix exclusively forwarded to itself if the host can be informed of the exclusivity. (In the case of a

DHCPv6 Prefix Delegation the prefix can be assumed to be of exclusive use by the requesting node, in accordance with the model in [RFC3633].) An implementation can, for example, safely "bind to an IPv6 subnet" in the style of <http://www.potaroo.net/ispcol/2016-09/subnetbind.html>, or start 64sharing [2] (given a prefix of a suitable size).

This memo documents an additional bit in the IPv6 RA PIO that makes this information explicit to receiving node.

3. Terminology

3.1. 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 RFC 2119 [RFC2119].

3.2. Abbreviations

Throughout this document the following terminology is used purely for the sake of brevity.

3.2.1. PIO-X

The term "PIO-X" is used to refer to a Prefix Information Option (PIO) that has the X bit set.

3.2.2. PIO-X RA

The phrase "PIO-X RA" is used to refer to an IPv6 Router Advertisement (RA) that contains one or more PIO-X entries (the same RA may also contain one or more PIOs without the X bit set).

3.2.3. Host

The term "host" may be used interchangeably throughout this document to mean a network node receiving and processing an RA. The receiving node may itself be a router, or may temporarily become one by routing all or a portion of an exclusive use prefix.

3.3. Concepts

Critical to correct network operations when employing PIO-X is the concept that both the router transmitting a PIO-X RA and its intended recipient be reasonably assured of the prefix's exclusivity. In support of this, the router must have confidence in the host's

presence and reachability, and several constraints are placed on the format of the RA for the host to validate.

3.3.1. Link layer guarantees

TODO: Add definition of things like "guaranteed point-to-point link" and what it is meant by having certain link-layer guarantees.

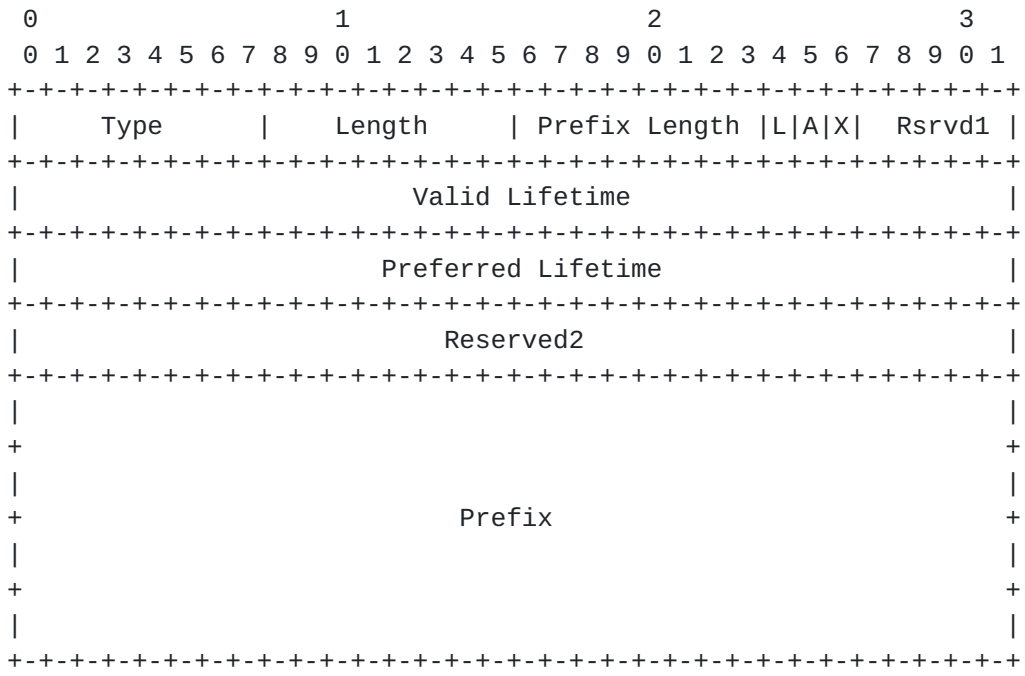
3.3.2. Likely sole intended recipient

TODO: section about "likely sole recipient" which can be referenced from other sections.

4. Updated Prefix Information Option

This document updates the Prefix Information Option specification in RFC 4861, section 4.6.2 with the definition of a bit from the former Reserved1 field as follows.

4.1. Updated format description



Fields:

X	The eXclusive use indicator flag, defined by this document. When set, the receiving node can be assured that all traffic destined to any address within the specified Prefix will be forwarded to itself by, at a minimum, the router from which the encapsulating RA was received, but possibly other routers as well. When not set, the receiving node MUST NOT make any assumptions of exclusive use of the specified Prefix, i.e. processing is unchanged from previous standards behavior.
Rsrvd1	Retains the same meaning as Reserved1 from <RFC4861> section 4.6.2.
All other fields	Retain their same meaning from <RFC4861> section 4.6.2.

4.2. Processing

Nodes compliant with this specification perform the following additional processing of RAs and PIO-X options when a PIO-X option is present.

4.2.1. Verify sole recipient

A node receiving a PIO-X option MUST verify that it is the (likely) sole intended recipient of the PIO-X RA. This done by verifying that the RA is unicast to the node at the IPv6 layer and, if applicable, at the link layer. On links that provide the node with a guarantee that it is the only possible PIO-X RA recipient (e.g. PPP, 3GPP links) this validation step SHOULD NOT be performed.

If an address other than :: (the unspecified address) was used as the source address for one or more Router Solicitations (RS) on this link, the node MUST verify that the IPv6 destination address of the PIO-X RA is one of the RS source addresses in use. If the link over which this communication takes place is known to be point-to-point, i.e. the nature of the link ensures that the node is the only possible recipient of an RA this check SHOULD NOT be performed.

If the node receives a PIO-X RA over a link-layer medium that supports link-layer addresses, it MUST verify that the link-layer destination address of the PIO-X RA is its own link-layer address. If the node received a PIO-X RA over a point-to-point medium (such as PPP) this step is unnecessary.

If any part of this "sole unicast recipient" verification fails, the node MUST ignore the PIO-X bit and continue processing as if it were not set (X=0).

4.2.2. (Re)Interpretation of other flags

Nodes compliant with this specification, i.e. those that understand the X-bit, MUST, when the X-bit is set, ignore the actual values of the L and A flags and instead interpret them as follows:

- interpret the L bit as if it were 0 (L=0)

- interpret the A bit as it it were 1 (A=1)

The rationale for this is as follows.

4.2.2.1. PIO L bit

Because a PIO-X aware node will know that it has exclusive use of a prefix with non-zero valid lifetime, the prefix itself cannot be considered to be on-link with respect to the link on which the PIO-X RA was received.

Note that a given address from within the prefix may be considered on-link according to the definition in <RFC5942> section 4, item 1, should the receiving node chose to configure that address on said link, but this is in no way synonymous with the entire prefix being considered on-link.

4.2.2.2. PIO A bit

Because a PIO-X aware node will know that it has exclusive use of a prefix with non-zero valid lifetime, autoconfiguration of addresses according to any desired scheme, e.g. <RFC4862>, <RFC7217>, et cetera, is implicit in the setting of the X bit.

Accordingly, the A bit can be interpreted as having been set, should the host choose to apply standard address generation schemes that require the bit to be set. It is free to assign any address formed from an exclusive prefix to any available interface; it is not required to configure the address on the link over which the PIO-X RA was received (i.e. it is under no obligation to form addresses such that they would be classified as on-link (according to the definition in <RFC5942> section 4, item 1).

4.3. Transmitting PIO-X RAs

When a router transmits an RA containing one or more PIO-X options it MUST unicast the PIO-X RA to its intended recipient at the IPv6 layer and, if applicable, at the link-layer.

It is RECOMMENDED that a PIO with the X-bit set also have the PIO flags L=0 and A=1 explicitly configured, for backward compatibility (i.e. use by non X-bit aware nodes).

5. Host behavior

TODO: This section needs some work.

5.1. PIO-X processing

A receiving node compliant with this document processes an RA with a PIO entry with the X flag set according the requirements in previous standards documents (chiefly <RFC4861> section 6.3.4) subject to the additional requirements documented in Section 4.2.

5.2. Neighbor Discovery implications

5.2.1. Duplicate Address Detection (DAD)

Whatever use the host makes of the exclusive prefix during its valid lifetime, it SHOULD NOT perform Duplicate Address Detection ("DAD", <RFC4862> section 5.4) on any address it configures from within the prefix if that address is configured on either the interface over which the PIO-X RA was received or on a loopback interface. Note that this does not absolve the host from performing DAD in all scenarios; if, for example, the host uses the prefix for 64sharing [3] it MUST at a minimum defend via DAD any addresses it has configured for itself as documented in Requirement 2 of <RFC7278> section 3.

5.2.2. Router Solicitations (RSes)

Routers announcing PIO-X RAs do so via IPv6 unicast to the intended receiving node and may note the IPv6 unicast destination address of an RS as the next hop for the exclusive prefix. As such, hosts compliant with this SHOULD NOT use the unspecified address (::) when sending RSes; they SHOULD prefer issuing Router Solicitations from a link-local address.

It is possible for a node to receive multiple RAs with a mix of exclusive and non-exclusive PIOs and even non-zero and zero default router lifetimes. While it is not possible for a host (receiving

node) to be sure it has received all the RA information available to it, hosts compliant with this specification SHOULD implement Packet-Loss Resiliency for Router Solicitations [RFC7559] so that the host continues to transmit Router Solicitations at least until an RA with a non-zero default router lifetime has been seen.

5.3. Link-local address behavior

Routers announcing PIO-X RAs may record the source (link-local) address of an RS as the next hop for the exclusive prefix. A node compliant with this specification MUST continue to respond to Neighbor Solicitations for the source address used to send RSes (alternatively: the destination address of unicast PIO-X RAs received). Hosts that deprecate or even remove this address may experience a loss of connectivity.

5.4. Source address selection

No change to existing source address selection behavior is required or specified by this document.

5.5. Next hop router selection

No change to existing next hop router selection behavior is required or specified by this document.

5.6. Implications for Detecting Network Attachment

TODO: Describe implications for Detecting Network Attachment in IPv6 [4] (DNAv6). Probably the best that can be done is (a) no change to RFC6059 coupled with (b) a host MAY send a test packet (e.g. ICMPv6 Echo Request) with a source and destination address from within the PIO-X prefix to the PIO-X RA issuing router and verify the packet is delivered back to itself. Consistent failure to receive such traffic MAY be considered a signal that the exclusive prefix should no longer be used by the host.

5.7. Additional guidance

The intent of networks that use PIO-X RAs is not to enable sophisticated routing architectures that could be far better handled by an actual routing protocol but rather to propagate a prefix's exclusive use information to enable the receiving node to make better use of the available addresses. As such:

A PIO-X receiving node SHOULD NOT issue ICMPv6 Redirects ([RFC4861] section 4.5) for any address within an exclusive use prefix via the link over which the PIO-X RA was received.

Redirecting portions of exclusive prefixes to other "upstream" on-link nodes is not a supported configuration.

A PIO-X receiving node SHOULD NOT transmit RAs with any subset of its exclusive prefixes via the same interface through which the exclusive prefix was learned.

6. Router behavior

TODO: This section needs some work.

6.1. PIO-X RA destination address

Since the host will not perform DAD for addresses within prefix announced via PIO-X, it's very important that only a single host receives the PIO-X RA. Therefore, the router MUST only include PIO-X in RAs that are sent using unicast RAs to destination unicast link-layer address and IPv6 link-local unicast address for a specific host. For point-to-point media without link-layer addresses or where there is guaranteed to only be single host that will receive the PIO-X RA (e.g. as enforced by link layer mechanisms), the router MAY send PIO-X RA with multicast destination IPv6 address. Under all circumstances the router MUST maintain a binding table of state information as discussed in Section 6.3.

6.2. Detecting hosts to send PIO-X RAs to

When the host starts using a network connection it normally sends out an RS (Router Solicitation) packet. This is one way for the router to detect that a new host is connected to the network and detects its link-local address. If the router is configured to use PIO-X, it can now perform necessary processing/configuration and then send the PIO-X RA.

For some networks, the host information regarding link-layer and link-local address might be available through other mechanism(s). Examples of this are PPP, 802.1x and 3GPP mobile networks. In that case this information MAY be used instead of relying on the host to send RS. It is however RECOMMENDED that these networks also provide indication whether the host is no longer connected to the network so that the router can invalidate the prefix binding prior to binding expiration (timeout).

6.3. Binding table requirements

Routers transmitting PIO-X RAs have state maintenance and operational requirements similar to delegating routers in networks where DHCPv6

Prefix Delegation [RFC3633] is used. The state maintained is describe here in terms of a conceptual binding table.

R1 The router SHOULD keep track of which PIO-X prefix has been issued to each node.

R2 The router SHOULD keep the binding between prefix and link-local address for the advertised valid lifetime, plus some operationally determined delay prior to reissuing a prefix ("grace period"), of the prefix.

R3 The router MUST monitor the reachability of each node in the binding table via Neighbor Unreachability Detection ("NUD", <RFC4861> section 7.3) or an equivalent link-layer mechanism.

R4 The binding SHOULD be considered refreshed every time a periodic PIO-X RA is sent to a node.

R5 If the router is informed by some other mechanism (link-layer indication for instance) that a node is no longer connected to the link, it MAY immediately invalidate the prefix binding. (DISCUSS: Is this the correct approach? Do we want to point to some definition somewhere else?)

6.4. Preparations before sending a PIO-X RA

When the router intends to send a PIO-X RA, it SHOULD before sending the PIO-X RA, complete any and all necessary processing for the host to start using the PIO-X prefix to communicate through the router to other networks. This is so that the host can start using PIO-X based addresses without delay or error after receipt of the PIO-X RA.

6.5. Implementation considerations

TODO: Out of scope things that are worth careful consideration include...

Routers SHOULD NOT announce the same prefix to two different nodes within the valid lifetime of the earlier of the two PIO-X announcements.

A link may operate in a mode where routers announce RAs to all nodes, possibly with non-exclusive PIO data, and non-zero default router lifetimes. Separately, one or more other nodes on the link may announce exclusive PIO information to nodes along with zero default router lifetimes. Except in the presence of a non-expired more specific route, e.g. learning from an <RFC4191> Route Information Option (RIO), the receiving node should send exclusive use prefix

originated or forwarded traffic destined off-link through routers with non-zero default router lifetimes.

7. Acknowledgements

8. IANA Considerations

This memo contains no requests of IANA.

9. Security Considerations

This document fundamentally introduces no new protocol or behavior substantively different from existing behavior on a link which guarantees a unique /64 prefix to every attached host. It only describes a mechanism to convey that topological reality, allowing the host to make certain optimizations as well as share the exclusive prefix as it sees fit with other nodes according to its capabilities and policies.

10. References

10.1. 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>.

[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>.

[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>.

[RFC7559] Krishnan, S., Anipko, D., and D. Thaler, "Packet-Loss Resiliency for Router Solicitations", RFC 7559, DOI 10.17487/RFC7559, May 2015, <http://www.rfc-editor.org/info/rfc7559>.

10.2. Informative References

[RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633, DOI 10.17487/RFC3633, December 2003, <http://www.rfc-editor.org/info/rfc3633>.

[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>.

[RFC4429] Moore, N., "Optimistic Duplicate Address Detection (DAD) for IPv6", RFC 4429, DOI 10.17487/RFC4429, April 2006, <http://www.rfc-editor.org/info/rfc4429>.

[RFC5452] Hubert, A. and R. van Mook, "Measures for Making DNS More Resilient against Forged Answers", RFC 5452, DOI 10.17487/RFC5452, January 2009, <http://www.rfc-editor.org/info/rfc5452>.

[RFC7934] Colitti, L., Cerf, V., Cheshire, S., and D. Schinazi, "Host Address Availability Recommendations", BCP 204, RFC 7934, DOI 10.17487/RFC7934, July 2016, <http://www.rfc-editor.org/info/rfc7934>.

10.3. URIs

[1] RFC4429

[2] RFC7278

[3] RFC7278

[4] RFC6059

Authors' Addresses

Erik Kline
Google Japan KK
6-10-1 Roppongi
Mori Tower, 44th floor
Minato, Tokyo 106-6126
JP

Email: ek@google.com

Mikael Abrahamsson
T-Systems Nordic
Kistagangen 26
Stockholm
SE

Email: Mikael.Abrahamsson@t-systems.se