                 IPv6 Prefix Assignment in Small Networks
                 draft-baker-homenet-prefix-assignment-01

Abstract

   It is necessary to allocate prefixes in small networks, which include
   residential and Small Office/Home Office (SOHO) networks in a manner
   that minimizes or eliminates manual configuration. This note
   suggests an approach.

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 September 9, 2012.

Copyright Notice

Table of Contents

1. Introduction

   One of the objectives of the design of IPv6 [RFC2460] has been to
   reduce or minimize the need for manual configuration in networks.
   IPv4 [RFC0791] networks, when it became widely deployed in the
   1980's, required manual configuration, and the scaling limits of the
   approach quickly became apparent. One of the outcomes of that was
   the Dynamic Host Configuration Protocol [RFC2131] (DHCP), which
   facilitated central administration of desktop computers. In
   practice, DHCP itself has been of limited utility in the
   administration of network equipment; while it is conceptually
   possible to use it for any kind of configuration, more flexible
   protocols such as the Network Configuration Protocol
   [RFC6241][RFC6242] have been preferred.

   Allocation of prefixes in small networks calls for an approach that
   can be completely automated. This note documents a procedure that
   has been suggested by several. It builds on a few basic assumptions:

   o IPv6 prefixes are allocated to a small network by one or more
      upstream service providers using [RFC3315] and [RFC3363].

   o IPv6 prefixes may allocated to LAN within a small network by the
      CPE Router using [RFC3315] and [RFC3363].

   o Occasional inefficiencies such as allocating two /64s to a LAN
      from a given upstream prefix are acceptable, especially if short-
      lived.

   o Small networks, such as described in Home Networking Architecture
      for IPv6 [I-D.chown-homenet-arch], are simple enough in structure
      that the mechanism described in this note is adequate.

   These assumptions bear analysis. The first two, that prefixes can
   and may be allocated using mechanisms designed for the purpose, seems
   self-evident. The third builds on the IPv6 premise that a host may
   have more than one prefix on an interface and one or more addresses
   in each prefix; in such a case, while it may be suboptimal to
   allocate more than one /64 from the same upstream prefix, the hosts
   will not complain and the routing protocols will route them. The
   fourth may be considered the limit of applicability; if a network
   requires a prefix aggregation design or is otherwise too complex for
   this procedure to be effective, other procedures are more
   appropriate.

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. Scope of this Document

   This document describes a procedure for prefix delegation and
   assignment. It results in the assignment of a series of /64 prefixes
   on the links in a small home network.

   While this document describes the use of DHCPv6 for prefix
   delegation, specification of the use of DHCPv6 for address assignment
   and other purposes is out of scope.

   If a network includes interior routers and the CPE router is not
   directly to all of the links in the network, the routers in the
   network will need routing information to forward traffic in the
   network and between the network and the service provider network.
   The specification of a routing protocol or other mechanism to provide
   that routing information to the routers is beyond the scope of this
   document.

3. Simple Tree Network Case

   The first case to describe is that of a network with a simple tree
   topology. In this network, there is a single CPE router attached to
   a single SP network. The interior of the network is organized as a
   tree. Each interior router has one "upstream" interface and one or
   more "downstream" interfaces. Each link in the network has a single
   interior router with a downstream interface attached and zero or more
   interior routers with an upstream interface attached.

   The fundamental procedure for prefix allocation takes three phases:

   o Allocating a prefix from the upstream network,

   o Prefix allocation by the CPE Router, and

   o Prefix allocation by a subsequent router.

3.1. Assignment of prefixes in a simple network

   This section describes the assignment of prefixes in a simple
   network. The network is assumed to be tree-structured, including one
   CPE router that is connected to a SP network and one or more interior
   routers. The interior routers each have a single "upstream"
   interface and one or more "downstream" interfaces. The upstream
   interface of each interior router is connected to a link in the
   network to which a downstream interface of a router closer to the CPE
   router is already connected.

   The CPE router obtains a delegated prefix for the entire home
   network, and manages prefix allocations for all of the interior
   routers. Each interior router uses DHCPv6 on its upstream interface
   to obtain delegated prefixes from the CPE router for each of the
   interior routers downstream interfaces.

3.1.1. CPE Router Behavior

   The CPE router obtains a delegated prefix from the SP provisioning
   system using [RFC3315] and [RFC3363] and other appropriate
   provisioning systems. The prefix delegated from the service provider
   includes a preferred and valid lifetime for the prefix.

   Once the CPE router has received a delegated prefix, it assigns a /64
   subprefix to each of the links to which the router is attached. The
   CPE router configures an address to each of its interfaces from the
   prefix assigned to the link to which the interface is attached.

   After assigning the interface addresses, the CPE router begins
   sending Router Advertisement (RA) messages [RFC4861] advertising the
   appropriate prefix on each attached link. The RAs include preferred
   and valid lifetimes derived from the lifetimes associated with the
   delegated prefix from the service provider. The RA also advertises
   the CPE router as the default router for the link. Other fields in
   the RAs are set as appropriate.

   At this point, the links to which the CPE router is attached are now
   provisioned with prefixes taken from the prefix obtained from the
   service provider. The CPE router uses ongoing DHCPv6 messages
   exchanges according to [RFC3315] and [RFC3363] to maintain and update
   its delegated prefix.

   The CPE router implements a DHCPv6 server for prefix subdelegation
   throughout the rest of the network. In preparation for assigning
   prefixes to links in the rest of the network, the CPE router makes
   all of the remaining prefixes from the network prefix available for
   subdelegation through a DHCPv6 server. The CPE router configures the

preferred and valid lifetimes for the subdelegated prefixes from the
values received from the service provider.

3.1.2. Interior Router Behavior

When an interior router is connected to the home network, its
upstream interface is attached to a link in the home network, and its
downstream interfaces are connected to other links to be added to the
home network.

3.1.2.1. Network with a Tree Topology

After the upstream interface is attached to a link, the interior
router listens for RAs on the upstream interface and configures the
upstream interface according to the information contained in the
received RAs.

When the interior router receives an RA, the router initiates a
DHCPv6 message exchange to obtain any needed prefixes from the prefix
managed by the allocating router. The interior router requests the
delegation of a separate /64 prefix for each of its downstream
interfaces. The DHCPv6 service in the home network delivers the
DHCPv6 traffic between the interior router and the CPE router.

The CPE router delegates the requested prefixes from the prefix
delegated to the network. The interior router then assigns a prefix
to each link attached to which a downstream interface is attached,
configures those downstream interfaces with addresses from the
assigned prefixes and begins sending RAs on the downstream
interfaces. The preferred and valid lifetimes for the advertised
prefix are derived from the lifetimes in the DHCPv6 delegation, and
the RAs advertise the interior router as the default router for the
link.

3.1.2.2. Non-tree Topologies

It is quite likely that real world deployments will violate the
assumption in the previous section that only one downstream interface
will be attached to each link in the home network. In this
situation, it is desirable that the link only be assigned one prefix
and, therefore, only one of the interior routers with a downstream
interface on the link be responsible for assigning a prefix and
sending RAs on the link.

To avoid duplicate address assignment, a router first listens for RAs
on the link attached to its downstream interface. If the router does
not receive an RA after listening for INTERVAL1 microfortnights, the
router assumes it is responsible for assigning a prefix to that link

   and initiates the DHCPv6 process for obtaining a delegated prefix.

   After the router determines it is responsible for the link attached
   to its downstream interface, it continues to listen for RAs from
   other routers on the link. If it receives an RA from another router,
   it deassigns its delegated prefix from the link, unconfigures any
   addresses assigned from that prefix and releases the delegated prefix
   to the CPE router using DHCPv6.

      Discussion: there is a race condition in this; two routers may
      make the decision to manage the link's prefix simultaneously. To
      avoid this, the timing should be jittered enough to make this
      unlikely.

   If a router hears an RA such as described in Section 3.1.2, it uses
   IPv6 Stateless Address Autoconfiguration [RFC4862][RFC4941] or a
   DHCPv6 [RFC3315] request to each announced allocator to generate an
   address within the prefix for use in that subnet.

   After the router determines that some other router is responsible for
   the link attached to its downstream interface, it continues to listen
   on the interface for RAs. If the router receives no RA on the
   interface for INTERVAL2 microfortnights, the router takes
   responsibility for the link and initiates the process described above
   to obtain and assign a prefix to the link.

3.1.2.3. Multi-homed Network

   If a network has multiple service provider networks, it will have
   multiple prefixes. This situation is easiest to describe if the
   network is connected to each service provider through a separate CPE
   router.

   Each CPE router obtains a delegated prefix from its service provider
   and then manages the prefix according to the discussion in Section 1.

   First layer of interior router get multiple direct DHCPv6 prefixes.
   Assigns each prefix in parallel. Sets up DHCPv6 relay agent to point
   to each of the CPE routers.

   Next layer receives DHCPv6 transaction from each CPE router because
   upstream router forwards DHCPv6 messages to each of the CPE routers.

4. Issues in a simple cascade procedure

   There are a number of potential issues in this procedure.

4.1. Sequence of subnet number allocation

Apart from cases in which the administration has chosen to fix a
given subnet to a given LAN, such as to support server deployment in
DNS, it is generally advised that subnet numbers be randomized. This
is to make certain network attacks a little more difficult.

4.2. Multihoming Issues

One issue is "what happens if one has multiple upstream networks with
multiple CPE Routers and therefore multiple allocators?" The design
of the RA information element announcing the allocator is intended to
simplify that by announcing an allocator.

4.3. Race Conditions

In the simplest case, there are no race conditions; the home has
exactly one router, it obtains a prefix from its upstream network,
and sub-allocates to its interfaces. If there are additional routers
in the home, however, either there are one or more links that are not
attached to the CPE Router or there are zero; in the event that there
are one or more such links, they may be connected by one router or by
multiple routers.

One race condition is when two interior routers are attached to the
same LANs as the CPE. For example, one might have a wireless router
in the home that connects both to the wired and the wireless network
that the CPE Router is on. In such a case, it will hear and
interpret one of the CPE Router's RAs first, and then the other some
amount of time later. The purpose of the INTERVAL1 delay in
Section 3.1.2 is to allow this race condition to stabilize before the
router acts on this information it has.

A second race condition occurs when two "subsequent" routers are on
the same LAN but it is not serviced by the CPE Router. These routers
will both use the procedure of Section 3.1.2 to attempt to allocate a
prefix to the LAN and so create a subnet. It is RECOMMENDED that the
allocator allocate at most one prefix per INTERVAL2, ignoring all
other requests, in order to allow the "subsequent" routers to sort
out this class of race condition. If needed, ignored routers will
re-request the allocation.

Due to the possibility of packet loss in the network, it is possible
that these race conditions may result in a given LAN developing
multiple subnets. While suboptimal, this is not a violation of the
architecture and should cause no issues. However, in the event that
two routers observe that they are announcing different subnets in the
same upstream prefix on the same LAN, the one with the numerically

least subnet number SHOULD NOT allow its prefix to expire, but any
others SHOULD allow their prefixes to expire.

4.4. Scaling Issues

Obviously, use of this procedure in a complex network results in a
serialization of prefix allocation that may take more time to settle
than is operationally desirable (number of LANs times INTERVAL2). In
such cases, the administration will have to decide how it wants to
handle the issue. One approach would be to divide the network into
easily-aggregated sections and use the procedure within each section;
another would be to use a different procedure.

In such networks, the routers requesting prefixes can also act as a
denial of service attack, by flooding the CPE Router with requests.
Given that the procedure eventually terminates, this is undesirable
but of limited duration.

4.5. Prefix Stability

In networks that contain servers or names that are announced in DNS,
it is often valuable to have the same LAN always have the same subnet
number applied to it. The procedure as described could accomplish
that if the CPE Router maintains memory of what router it has
allocated a given prefix to recently, or would fail to provide that
if it does not. The distinction is essentially a marketing
requirement that the implementation will need to decide for itself.

4.6. When you run out of prefixes

If a network runs out of subnet numbers and therefore subnet
prefixes, this is considered a provisioning failure. It can result
when multiple prefixes are allocated to the same LAN, which should be
unusual and will end when one of the routers releases its prefix. It
can also result when the upstream network allocates a prefix that is
too long and as a result contains too few potential prefixes. In
that case, the administration is forced to either reorganize its
network or negotiate for a shorter prefix.

5. IANA Considerations

This specification makes no request of the IANA.

6. Security Considerations

<TBD>

6.1. Privacy Considerations

   <TBD>

7. Change Log

   Initial Version: 4 October 2011

   March 2012 Removed RA option.

8. References

8.1. Normative References

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

8.2. Informative References

   [I-D.chown-homenet-arch]
              Arkko, J., Chown, T., Weil, J., and O. Troan, "Home
              Networking Architecture for IPv6",
              draft-chown-homenet-arch-01 (work in progress),
              October 2011.

   [RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791,
              September 1981.

   [RFC2131] Droms, R., "Dynamic Host Configuration Protocol",
              RFC 2131, March 1997.

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

   [RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C.,
              and M. Carney, "Dynamic Host Configuration Protocol for
              IPv6 (DHCPv6)", RFC 3315, July 2003.

   [RFC3363] Bush, R., Durand, A., Fink, B., Gudmundsson, O., and T.
              Hain, "Representing Internet Protocol version 6 (IPv6)
              Addresses in the Domain Name System (DNS)", RFC 3363,
              August 2002.

   [RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman,
              "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861,
              September 2007.

   [RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless
              Address Autoconfiguration", RFC 4862, September 2007.

   [RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy
              Extensions for Stateless Address Autoconfiguration in
              IPv6", RFC 4941, September 2007.

   [RFC6241] Enns, R., Bjorklund, M., Schoenwaelder, J., and A.
              Bierman, "Network Configuration Protocol (NETCONF)",
              RFC 6241, June 2011.

   [RFC6242] Wasserman, M., "Using the NETCONF Protocol over Secure
              Shell (SSH)", RFC 6242, June 2011.

Authors' Addresses

   Fred Baker
   Cisco Systems
   Santa Barbara, California 93117
   USA

   Email: fred@cisco.com

   Ralph Droms
   Cisco Systems
   1414 Massachusetts Avenue
   Boxborough, Massachusetts 01719
   USA

   Email: rdroms@cisco.com