Best Current Operational Practice for operators: IPv6 prefix assignment for end-users - static (stable) or dynamic (non-stable) and what size to choose.

DRAFT v.1 (2017-03-27)

Authors: Jan Žorž <zorz@isoc.org>, **Sander Steffann** <sander@steffann.nl>, **Primož Dražumerič** <Primoz.Drazumeric@telekom.si>, **Mark Townsley** <townsley@cisco.com>, **Andrew Alston** <andrew.alston@liquidtelecom.com>, **Gert Doering** <gert@space.net>, **Jordi Palet** <jordi.palet@consulintel.es>, **Jen Linkova** <furry@google.com>, **Luis Balbinot** <lbalbinot@brfibra.com.br>, **Kevin Meynell** <meynell@isoc.org>
Working Groups: RIPE BCOP TF and RIPE IPv6 WG

Table of content:

1. What is a BCOP?

A Best Current Operational Practices (BCOP) document describes best current operational practice on a particular topic, as agreed by subject matter experts and periodically reviewed by the Internet community. This document is not intended to document what practices may be in future and what they might look like, but to reflect the best methods of implementing IPv6 at the time of publication. Updates to this document will be published to reflect changes in best current practices where there are developments in standards and implementations.

2. Introduction and incentives

The task of ISPs is to provide their customers with Internet connectivity, and that includes providing those customers with IP addresses. Customers generally receive a single IPv4 address, but with IPv6 a customer requires a block of addresses, also known as a prefix. This document provides guidance as to what size such prefixes should be.

Network operators may incur surprises if they make a wrong decision when choosing the size of a prefix assignment, or the method of assigning it. This may be immediate in terms of other networks or content providers measuring the IPv6 “brokenness” who may stop serving you AAAA records, or it may happen in the future when, for example, the HOMENET standard results in more complex home networks being built and operated. In these cases, network operators will have to adapt their IPv6 implementations to new requirements, possibly including complete network renumbering, whilst their IPv6 network is already in production and has connected users.

Up until now, there have been no clear recommendations as to what size of prefixes to choose and assign to customers, hence the decision to document some best current operational practices for network operators to consider when making their decisions.

3. Size of end-user prefix assignment: /48, /56 or something else?

Network operators deploying IPv6 will sooner or later need to decide which sizes of IPv6 prefixes they will assign to their end customers. At this point, the IPv4 mentality and perception of scarcity of numbering resources needs to be reconsidered because the abundance of IPv6 addresses enables numbering decisions to be taken differently.

Many operators might perceive that assigning large IPv6 prefixes to end customers is wasteful, but the reality is that decisions should be based on the IPv6 protocol architecture design. For example, Tony Hain calculated that assigning a /48 to every human on Earth, and never recovering those, will still mean that IPv6 would have a lifetime over the 480 years. On that timescale there will be other reasons, not just scarcity of IPv6 addresses that will require the IETF to design a successor to IPv6.

3.1. Numbering the WAN link (interconnection between our network and the end-user CPE):

Before going into details about the size of IPv6 prefix assignments, the choice for the WAN link needs to be understood. There are three options for addresses on the link between the operator network and end-user CPE WAN port.

3.1.1. /64 prefix out of the IPv6 prefix assigned to the end-user

Some operators use the first or last /64 prefix to number the CPE WAN link, from the larger IPv6 prefix assigned to the end customer. This has been described by and old IETF Internet Draft (https://tools.ietf.org/html/draft-palet-v6ops-point2point-01).

This is sometimes a good idea, but in some cases may fail as some CPEs or scripts may try to use it for loopback and assign the next /64 to the LAN ports. Responsibility for the prefix is delegated to a CPE, but then one /64 is used for the WAN connection, and whilst there is a DHCPv6 option to let the CPE know if one of its prefixes have been “stolen” (RFC 6603), not all CPEs support this and some issues will be encountered.

This is an option that simplifies routing and general provisioning provided you’re sure that the CPEs in your network don’t have the aforementioned issues.

3.1.2. /64 prefix out of a dedicated pool of IPv6 prefixes

Using a /64 prefix from a dedicated pool of IPv6 prefixes is the most common scenario and currently the best practice. A separate block of IPv6 space is allocated for the WAN links to the end customer CPEs, so that when CPE connects to the network and performs router discovery, a /64 prefix is used to number both ends of the connection.

In the case of an end customer host (not router) connecting over PPPoE (for example), the set-up process is concluded when the host is properly IPv6 numbered and can start sending and receiving traffic.

In the case of a router connecting to a network that is capable of issuing a DHCPv6 Prefix Delegation (PD) request, the IPv6 prefix delegation process is started and an IPv6 prefix is assigned to the CPE. This method is most probably the safer one, covering most of the scenarios regardless of underlying technologies, and is the one most used by operators.

Some operators also use /126, /127, /112 or an alternative prefix assignment instead of /64 for managed links where addresses on the link are manually configured, and according to RFC 6164, a /127 is suggested. Please note, that not all equipment currently supports this option, so /64 is still the safest approach and has the advantage of being future proof in the event that new standards make usage of the other 64 bits for other purposes or the link becomes point-to-multipoint, etc.. Furthermore, some ISP hardware has limitations in using prefixes longer than /64, so the use of non-/64 prefixes will use more resources on these devices. If we decide to use /127 for each point to point link, then it is also advisable to allocate a /64 for each link and use just one /127 out of it to prevent the Neighbor Discovery exhaustion attack.

3.1.3. Unnumbered

Some operators decide to leave the WAN link to the end-user CPE unnumbered which means no GUA (IPv6 Global Unicast Addresses), so only link-local addresses are used at both ends of the link. While this might work for routers, it does not work for devices that cannot request a prefix delegation over DHCPv6 and are therefore left without any usable GUA to allow traffic in or out. In the case of a router on the end customer side, the route for the assigned prefix is pointed towards the link-local address on the CPE/router WAN port. and the default route on the CPE/router is pointed towards the link-local address on the upstream network equipment port. This method may be seen as easier to implement, but it also brings some drawbacks such as difficulties with troubleshooting and monitoring that need to be considered.

It is most useful in scenarios where it is known that only CPEs will be attached to the WAN link, and where a pingable address of the CPE is known (e.g. because the ISP-provided CPEs are always pingable on the first delegated address). It is less suitable in situations where unknown CPEs and/or non-CPE devices are attached to the WAN link.

In addition, non-routers (e.g. a Windows/Linux PC or any other end-user host) connecting to a network that have no ability to ask for prefix delegation over DHCPv6 might experience problems and will stay unnumbered upon connection if a /64 prefix is not used to number the CPE WAN link.

3.1.4. ULA

Some operators, use ULAs (IPv6 Unique Local unicast Addresses) for numbering the WAN link to the end-user CPE. This approach may cause numerous problems and is therefore **strongly discouraged**. For example, if the CPE needs to send an ICMPv6 message to a host outside an operator's network (to the Internet), the packet with ULA source address will not get very far and Path MTU discovery (PMTUD) will break.

3.1.5. Summary

Using a global /64 prefix for the WAN connection is the recommended choice.

Besides being a safe choice, using a /64 is sometimes required when there are more devices than the two endpoints on a WAN link (e.g. intermediary bridges or repeaters) that require management, or if there is the need for redundancy (e.g. VRRPv3 or multiple routers at the customer premises).

When choosing which /64 to use, the recommended option is to dedicate a separate pool of prefixes for the WAN links. While the addressing plan and administration might be easier when selecting the /64 from the prefix delegated to the customer, this is technically “stealing” because the customer’s CPE has been informed that whole prefix will be delegated to it, so it should not also be used on the WAN link unless it is known that all CPEs will support RFC 6603 to negotiate this.

Using ULA addresses on the WAN link is very strongly discouraged.

3.2. Prefix assignment options

To keep addressing plans usable and understandable, and to align with DNS reverse zone delegations, the size of the delegated prefix should align with a nibble boundary. Each hexadecimal character in an IPv6 prefix represents one nibble, which is 4 bits. The length of a delegated prefix should therefore always be a multiple of 4.

A single network at a customer site will be a /64. RIR policies permit assignment of a /48 per site, so the possible options for choosing a prefix size to delegate are /48, /52, /56, /60 and /64.

The following sections explain why /48 and /56 are the recommended prefix assignment sizes for end customers.

3.2.1. /48 for business customers and /56 for residential customers

Some operators decide to give a /48 prefix to their business customers and a /56 to their residential customers. This rationale is understood to be mainly coming from sales and marketing departments where they wish to create some distinction in services between different types of customer. This method can be considered as pragmatic, future-proof and has nearly no downsides, the same as “/48 for everyone approach”.

For more advanced home users that are manually configuring their servers and routers, it also needs to be considered that local addressing is a bit harder for /56 prefixes, as you are unable to use all 4 characters between the double colons in the address notation, which means that users can make mistakes whilst making their addressing plans and sub-assign address to their devices outside their assigned prefix. This should not be a problem in simple cases where a CPE assigns all prefixes, but it might cause annoyances for more advanced users that want to offer content and services from their home network.

It should be remembered that some mechanisms use a default /48 prefix size, so if a user already has an addressing plan for that, they will be forced to redo it and renumber the same as if they were coming from another ISP. Again, this typically applies to advanced home users only.

Example of assigned prefixes out of 2001:db8::/40 :

2001:db8:aaaa:1a00::/56
2001:db8:aaaa:1b00::/56
2001:db8:aaaa:1c00::/56
2001:db8:aaaa:1d00::/56

For the first prefix, **2001:db8:aaaa:1aXX::/56**, the XX is the part that the advanced end-user can use to make their own addressing plan.

An alternative is to reserve a /48 for residential customers, but actually assign them just the first /56. If subsequently required, they can then be upgraded to the required prefix size without the need to renumber, or the spare prefixes can be used for new customers if it is not possible to obtain a new allocation from your RIR (which should not happen according to current IPv6 policies).

3.2.2. /48 for everybody

This is probably the most practical and most pragmatic way to assign IPv6 prefixes to end customer CPE devices, as in this case everyone has a /48 prefix and advanced end-users are less likely to make less mistakes when addressing their networks and devices, resulting in much less call centre time to sort out problems. It also has the advantage of sharing the same prefix size as ULAs and some transitional mechanisms, so this facilitates a direct mapping of existing customer addressing plans to the prefix delegated to them.

Example of assigned prefixes out of 2001:db8::/40 :
2001:db8:aaaa::/48
2001:db8:aaab::/48
2001:db8:aaac::/48
2001:db8:aaad::/48

For the first prefix, **2001:db8:aaaa:XXXX::/48**, the XXXX is the part that the end-user can use to make their own addressing plan. This is also the structure used in most addressing plan manuals and guides.

3.2.3. Less than /56

It is not recommended to assign prefixes smaller than /56 unless there are very strong and unsolvable technical reasons for doing this.

There are enough IPv6 addresses to delegate end customers a /48, so a /56 already represents a sizeable restriction. There is no need delegate fewer addresses than that, so if your IPv6 allocation is not sufficient to provide a /56 to each end customer (remember there are over 100 million /56s in a /29!), explain to your RIR that your initial allocation was too small and that you require a larger allocation based on your IPv6 implementation plan. Offering less than /56 can be feasible if just a /29 is allocated and 6RD or a similar transition mechanism is being used where an IPv4 address is embedded in the IPv6 prefix. However, even in this case a larger allocation from your RIR can be justified.

Assigning a /64 or smaller prefix is highly discouraged, does not conform to IPv6 standards, and will break functionality in customer LANs. With a single /64 the end customer CPE will have just one possible network on the LAN side and it will not be possible to subnetting, VLANs, alternative SSIDs, or have several chained routers in the same customer network etc…

Some CPEs use a /64 for the loopback interface and some may have have multiple LAN segments predefined (for example Guest WiFi network and LAN), so as soon as there is more than one LAN segment behind the CPE, exceptions will have to be added to the ISP provisioning system that will greatly complicate management. It is not possible to assign less than a /64 to each LAN port/segment/subnet/VLAN due to IPv6 protocol requirements (SLAAC, ND, etc..) that reserve the last 64 bits of an IPv6 address for the host.

Moreover, there is work in progress at IETF showing that it can be convenient in some cases to assign a /64 to each host (e.g. for security, isolation of customers, or having many virtual machines on single host), again explaining the need for delegating many /64s to each customer.

A growing number of operators are also using prefix colouring to deploy products with distinct Service Level Agreements (e.g. voice, data, video) and this requires at least a unique /64 for each product or service. If combined with home networking technologies under development, the number of prefixes increases quite quickly and also delegating multiple IPv6 prefixes at the same time will take place, so assignments less than /56 will probably require renumbering in the near future.

There is a clear exception to this rule when assigning addresses in a cellular network. In this case a /64 will need to be provided for each PDP context for cellular phones, whereas for LTE modems/routers it will still be necessary to choose a /48 or /56, in accordance with the aforementioned considerations.

4. End-user IPv6 prefix assignment: Static, dynamic (non-stable) or stable.

After determining the size of the IPv6 prefixes to be assigned to end customers, it is necessary to establish how to actually assign them. There are two methods: static or dynamic assignment.

Static assignment means that a prefix is assigned to a customer (typically an AAA) and that prefix will always stay the same regardless of how many times the customer (re)connects or renews the lease. It is static/stable for the same customer in the same link/location regardless of the provisioning system being used.

Dynamic assignment is where a bigger pool of IPv6 space is assigned to each termination point, so and as customers connect, they are randomly assigned prefixes from this when they (re)connect or the lease time expires. They may get the same prefix, but typically will be a different one (non-stable) unless the lease time is so high that it effectively become stable.

In this case, a customer will get the same prefix even if they switch off their router for several months.

4.1. Why dynamic (non-stable) assignments are easier than static ones.

When faced with the task how to deploy IPv6, it is easy to go for option that initially requires less effort, time and energy. However, this will usually create more problems later on.

The easiest way method is to assign prefixes from a pool of IPv6 addresses to termination points (BNGs or other equipment, depending on the type access network) and let the provisioning system decide what customers will get when they connect and ask for a prefix delegation over DHCPv6. This practice is carried over from IPv4 world where addresses are commonly assigned dynamically and where CPEs have a NAT mechanism to conserve IPv4 space. Bear in mind that that end customers with an IPv4 subnet behind their CPE never got dynamically assigned IPv4 prefixes as this would require reconfiguration of all hosts on their network every few hours or days. By contrast, every IPv6 customer gets an IPv6 subnet so it is unnecessary to apply this IPv4 model to IPv6

Dynamic prefix assignment also appears initially easier as it facilitates aggregation of internal routing tables according to end customer connection termination points. Every termination point has its own pool of IPv6 addresses that are nicely aggregatable, whilst with static IPv6 prefix assignments it is necessary to discover which customer is terminated at which termination point, group them into larger IPv6 pools, and then update our database accordingly.

Some operators may ask what happens if a customer moves to another city or neighbourhood, but if they move to different location then all the equipment is switched off whilst transporting it. Since it will be necessary to provision the new connection elsewhere, then it is perfectly fine to also change the prefix of the customer to match the criteria for that aggregation point unless is in the same “aggregation region/point” or the customer specifically requests that the same IPv6 prefix moves with them. In this case, the internal routing table will be one entry bigger and the decision may be to offer that service at an extra cost, but that should be an uncommon scenario.

4.2. Why dynamic (non-stable) assignments are considered harmful.

Taking a scenario where there is a connected CPE with a dynamically assigned prefix (e.g. 2001:db8:aaaa::/48). The CPE may sub-assign local loopback addresses out of the first /64 segment (2001:db8:aaaa:0::/64) and sub-assign 2001:db8:aaaa:1::/64 to the first LAN interface (and possibly 2001:db8:aaaa:2::/64 to the guest WiFi interface, and so on).
So, the hosts on the network autoconfigure IPv6 addresses out of 2001:db8:aaaa:1::/64 and/or 2001:db8:aaaa:2::/64 segments and start communicating with the rest of the Internet.

Now all of a sudden there is a power outage (which is very common in many regions/countries in the world, even “highly-developed” countries) or the CPE freezes and reboots and the connection has to be established again, but this time with a new IPv6 assigned prefix of 2001:db8:bbbb::/48. If the CPE knows that the delegated prefix has changed it should send out RA packets with a prefix valid lifetime of 0 to tell all devices that the old addresses are no longer valid. However, the CPE rarely knows that before the reboot there was a different prefix on the network, and the packets to revoke the old IPv6 addresses do not get sent. In this case, multiple IPv6 addresses from completely different assigned prefixes end up on the same network interface, some of which will no longer work.

Different OS vendors treat this scenario differently, but there often ends up being a wrong source IPv6 address for sending packets through the CPE out to the Internet. As the network operator’s equipment deleted the previous delegated prefix route back to the CPE, any return packets never reach the originating device and IPv6 connectivity will be broken until the old IPv6 addresses time-out and are automatically removed from the interfaces. If another reconnection to the ISP is required in the interim, there will be a third set of IPv6 addresses from yet another assigned prefix, and this will cause even more confusion.

Some big content providers are measuring “IPv6 brokenness” in operator networks by matching the SYN, SYN ACK and ACK packets received/sent to/from a single source. If they see a SYN and send back SYN ACK, but never receive ACK in a timely manner, they slowly stop serving AAAA records for the AS number where they see this. So if you decide to assign IPv6 prefixes dynamically, don’t be surprised if content providers start ignoring your IPv6 traffic quite quickly and force your visitors back to IPv4. And these are usually destinations where you will be sending a significant amount of traffic.

Using dynamic prefixes also means that it is necessary to have a logging system that registers which WAN and the LAN prefixes are being used at each moment by each customer site.

Finally, if users have services such as web, email or VPNs, they typically need to manually configure the addresses of those, so using dynamic prefixes is not an option in this case. With the growth of broadband capacities (such as FTTH), it is becoming more and more common that end-users run servers or services on their LANs, including the likes of IP cameras or security systems in a home. Static/stable addresses will allow assign DNS names to be assigned to them.

If you wish to do dynamic (non-stable) prefix delegation, you must verify that the CPEs used by the majority of your users will not have the problems described above. If this can’t be ensured, then the recommendation is to avoid using dynamic non-stable prefixes and revisit the CPE and end user device implementations periodically to see whether this has become feasible.

It also needs to be realised that dynamic (non-stable) prefixes have other implications, because it is not only the CPEs that need to be numbered, but all the devices behind them in the customer LANs as well. Different implementations can have very different behaviors that may affect the number of support calls to your call centre every time a prefix changes.

If you wish to do dynamic prefix delegation today, verify with the CPE vendors used by the majority of your users that there will be no problems such as the ones described in this section. If this cannot be ensured, the recommendation today is to use stable (either static or long-lived stable dynamic assignments) prefixes today, and revisit the CPE and end user device implementations every few years to see if moving to fully dynamic assignments is becoming feasible for you.

4.3. Why static or stable prefix assignments are recommended.

As mentioned, the best practice in IPv4 was to assign static IPv4 prefixes whilst giving an end-user a routed prefix, and same is true for IPv6 where static prefix assignment is strongly recommended. If required for provisioning, the connection between a network and a customer CPE can be dynamically numbered using /64 (or even /127 if the equipment supports it), but choosing static IPv6 assigned prefixes for end customer LANs will avoid a lot of difficulties experienced by some earlier IPv6 network operators.

This is especially the case where there are non-residential customers as well, as this means you can have a single provisioning system and it is unnecessary to maintain a logging system (because each user/site always have the same prefixes and can be identified in the AAA) which reduces complexity and makes things cheaper. A bit of initial thought, planning and consideration for future needs can therefore save a great deal of time and energy when IPv6 has been deployed.

One other important consideration with static/stable address is that it possible to ‘name’ value added services using the DNS (e.g. camera1.username.ispname.com), which can result in considerable new income streams. Trying to deploy new services or applications with dynamic addresses is always more difficult and costly, and will increase time spent on troubleshooting.