

Date: 18 October, 2011

Shepard: Chris Grundemann

Subject Matter Expert(s) (SME): Chris Grundemann, Aaron Hughes, Owen DeLong

Status: BCOP

*BCOP Subject: IPv6 Subnetting*

*The content here within is intended to be original content authored by the Global Network Engineering Community (GNEC) “at-large” in an organized, democratic, “bottoms-up” approach.*

## Best Current Operational Practice (BCOP)

### 1. BCOP Summary (Appeal)

This BCOP aims to provide general IPv6 subnetting guidelines that can be followed by any network operator when planning and implementing an IPv6 network deployment. The primary focus is on understanding IPv6 subnets and addressing plans, creating operational clarity and future proofing.

### 2. BCOP Background / History

Due primarily to the lack of a need for the scarcity mentality inherent in IPv4 address planning, IPv6 subnetting brings with it a new paradigm that needs to be understood by network operators looking to roll out IPv6 networks, either purely greenfield or alongside current IPv4 networks (new logical networks sharing common infrastructure). This paradigm shift has prompted the Global Network Engineering Community (GNEC) to repeatedly request that the BCOP for IPv6 subnetting be defined and documented.

### 3. BCOP

#### 1. Every individual network segment requires at a minimum, one /64 prefix

A /64 is an IPv6 subnet which provides 64 network bits and 64 host bits. Regardless of the number of hosts on an individual LAN or WAN segment, every multi-access network (non-point-to-point) requires at least one /64 prefix. The IPv6 Addressing Architecture ([RFC 4291](#))<sup>1</sup> requires that all unicast addresses (that do not begin in binary 000) use a 64 bit Interface Identifier<sup>2</sup> constructed in Modified EUI-64 format.<sup>3</sup> This has ensured that the /64 prefix become the de-facto standard minimum prefix size. Additionally, this is a normative requirement of Stateless Address Auto Configuration (SLAAC, [RFC 4862](#))<sup>4</sup> and any network expected to perform SLAAC must have a /64 prefix available to it.

---

<sup>1</sup> Hinden, R., and S. Deering, “IP Version 6 Addressing Architecture”, RFC 4291, February 2006.

<sup>2</sup> See RFC 4291, Section 2.5.4. Global Unicast Addresses.

<sup>3</sup> For details see RFC 4291, Appendix A: Creating Modified EUI-64 Format Interface Identifiers.

<sup>4</sup> Thomson, S., Narten, T., and T. Jinmei, “IPv6 Stateless Address Autoconfiguration”, RFC 4862, September 2007.

While there are scenarios where it may be possible, or even advisable, to avoid this requirement; these are considered corner cases and not the general best practice, with one notable exception: Point-to-point links, detailed below.

## 2. Only subnet on nibble boundaries

A nibble boundary is a network mask which aligns on a 4-bit boundary. Nibble boundaries should be used to provide more human-readable addressing. This improves operational efficiency and reduces mistaken configuration by making the individual prefixes more recognizable and easier to understand sequentially. This is illustrated with several examples in Figure 1 below. As you can see in this example, the nibble aligned prefixes count upwards in a very predictable and obvious manner, the non-nibble aligned prefixes provide a much less apparent and easy to predict pattern. The latter often results in a higher tendency towards potentially costly “fat finger” mistakes.

Nibble Aligned Prefixes		Prefixes Not Nibble Aligned	
<u>2001:db8::/32</u>	<u>2001:db8::/32</u>	<u>2001:db8::/32</u>	<u>2001:db8::/32</u>
2001:db8::/36	2001:db8::/40	2001:db8::/37	2001:db8::/41
2001:db8:1000::/36	2001:db8:100::/40	2001:db8:800::/37	2001:db8:80::/41
2001:db8:2000::/36	2001:db8:200::/40	2001:db8:1000::/37	2001:db8:100::/41
2001:db8:3000::/36	2001:db8:300::/40	2001:db8:1800::/37	2001:db8:180::/41
<snip>	<snip>	<snip>	<snip>
2001:db8:c000::/36	2001:db8:9200::/40	2001:db8:6800::/37	2001:db8:8000::/41
2001:db8:d000::/36	2001:db8:9300::/40	2001:db8:7000::/37	2001:db8:8080::/41
2001:db8:e000::/36	2001:db8:9400::/40	2001:db8:7800::/37	2001:db8:8100::/41
2001:db8:f000::/36	2001:db8:9500::/40	2001:db8:8000::/37	2001:db8:8180::/41

Figure 1 - Examples of Nibble and Non-Nibble Boundaries

Note: Since the smallest subnet mask used for multi-access networks is a /64, simply counting down by 4 will give you the nibble boundary masks; 64, 60, 56, 52, 48, 44, etc.

Large networks that may grow beyond their initial allocation, and thus require a subsequent allocation, may want to consider not rounding up to nibble boundaries initially. This is especially true for networks using non-ARIN space.

Rounding up to the nearest nibble boundary represents an over subscription value of up to 16 times. Justification for additional allocations or assignments will require the demonstration of efficient utilization of the currently held space. Such a large over-sizing factor may result in some areas of the network having no space available while the overall utilization is too low to justify additional space.

This is not an issue in the ARIN region where rounding up to nibble boundaries is specifically allowed for, and a network is considered out of space when that network does not have available space to replenish any

*single pool within the addressing hierarchy (regardless of the utilization in other pools). However, this policy is not yet currently supported in the other regions. As such, networks that are likely to require a subsequent allocation from RIRs other than ARIN may want to consider rounding up only to standard CIDR boundaries (powers of two).*

### 3. Implement a hierarchical addressing plan to allow for aggregation

In a typical network, this equates to at least three levels of hierarchy; site, region and AS. A site can be a PoP, a building, a floor, or any other logical layer3 aggregation point within your network. **In general, each site should receive one /48.**

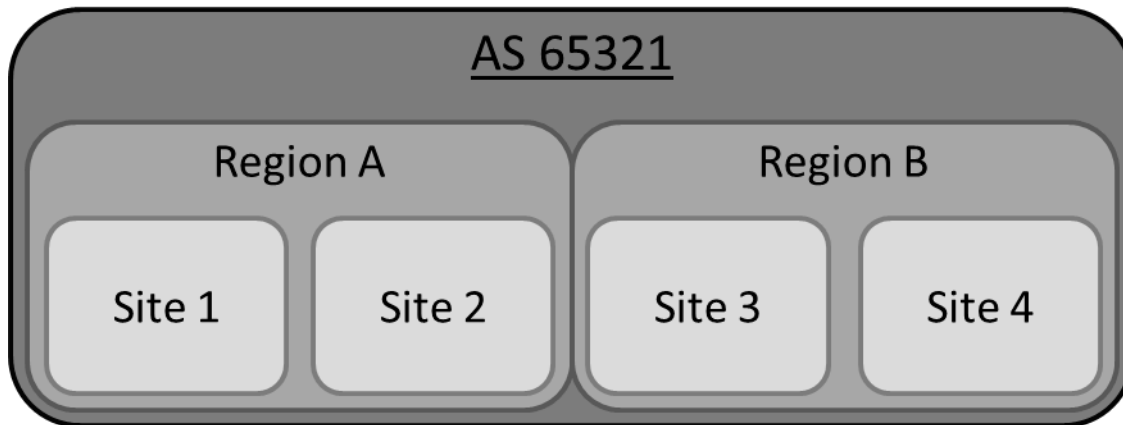


Figure 2 - Example 3-level Hierarchy

In some cases, a network may need additional aggregation points and thus a deeper hierarchy. In an extremely large network, this could be, for example; department, floor, building, campus, region, AS.

Another important consideration is that this large address space facilitates grouping types of usage within your address space (using common bit patterns for common uses). Grouping customer space (used by external parties), infrastructure space (interface addresses), internal space (used for the corporate network) critical internal space (RADIUS, TACACS, jump servers), etc., may make sense in your environment. Having a well-defined addressing plan that distinguishes between the various address usage types can simplify and reduce filter size.

### 4. One /48 from each region should be reserved for infrastructure

Reserve the first or last /48 from each region for infrastructure needs; interface numbering, loopback addresses, etc.

#### a. Loopbacks should be allocated from the first /64

The first (least specific, e.g. 2001:db8:100::/64 from 2001:db8:100::/48) /64 prefix of each region's infrastructure /48 should be reserved for loopback addresses. Loopbacks are only required to have a /128 mask (a single address) and thus can be numbered sequentially from the reserved /64.

Note well that IPv6 addresses are represented in hexadecimal, not decimal, notation. This means that "a" comes after "9", not "10". Of course you can skip hex digits and give your addresses the appearance of decimal numbering, which may be useful e.g. when matching loopback addresses to "router number" (i.e. although ::9 and ::10 are not sequential, that does not necessarily preclude you from numbering that way).

An alternative method is to allocate a full /64 for each loopback address, as described below for PtP links. This avoids violating RFC 4291's Interface ID requirement and it can be said that it isn't necessary to save IPv6 address space by using /128s on loopbacks. For example, out of a /48, if you reserved 16,384 /64s for loopbacks, you would be left with an additional 49,152 /64 subnets for links.

**b. Point-to-point links should be allocated a /64 and configured with a /126 or /127**

Allocating an entire /64 to each PtP link provides much more human readable addresses (e.g. X::1 and X::2). However, despite the requirement for a 64bit Interface ID mentioned above, a /126 or /127 mask should be used when configuring the actual network gear to avoid the Ping Pong, Neighbor Cache Exhaustion, and other similar issues, as documented in [section 5 of RFC 6164](#).<sup>5</sup>

Note well that there may be problems with using /127 prefix lengths in operation due to vendor implementation variance with regard to [subnet-router anycast](#).<sup>6</sup> This issue is documented in [RFC 3627](#),<sup>7</sup> specifically in section 3. Thorough testing in your own network environment is highly encouraged before attempting to use /127 prefixes in production, particularly in multi-vendor scenarios with any requirement for interoperability.

It is also worth noting that two addresses that are in the same /126 subnet will not be in the same /127 subnet, and vice versa. So /126 addressing might use X::1 and X::2, but with /127 subnets those would be in different networks, so you would have to use X:: and X::1, or X::2 and X::3.

**5. Sites/PoPs/locations and regions, etc. should be laid out such that within each level of the hierarchy, each subnet prefix is of equal size**

Within your hierarchical addressing plan, as described above, you should ensure that each unit at an equal level receives an equal sized prefix (i.e. each site should receive the same prefix size (e.g. /48) and each region should receive the same prefix size (e.g. /36)). This is once again a valuable aid in simplifying

---

<sup>5</sup> Kohno, M., Nitzan, B., Bush, R., Matsuzaki, Y., Colitti, L., and T. Narten, "Using 127-Bit IPv6 Prefixes on Inter-Router Links", RFC 6164, April 2011.

<sup>6</sup> See RFC 4291, Section 2.6.1. Required Anycast Address.

<sup>7</sup> P. Savola, "Use of /127 Prefix Length Between Routers Considered Harmful", RFC 3627, September 2003.

network configuration, operation, and troubleshooting. The vast increase in address space available to us with IPv6 allows for the creation of a homogenous, and therefore much more predictable, addressing plan.

The best way to ensure this is to plan based on the largest unit at each level. For example:

1. Take the largest site in your network, project its 5 year growth needs and allocate the required prefix for each of your sites.
2. Then look at your largest region, and allocate a prefix adequate to serve its needs for each of your regions.

Below are two examples to further illustrate this practice.

*In some networks there may be individual regions or sites that are so disparate in size or growth rate that this practice would result in very low (<50%) prefix utilization. This situation should be avoided through network design wherever possible. In extreme cases which cannot be avoided, it is recommended that operators take a “best fit” approach and simply allocate multiple prefixes to the largest or fastest growing sites or regions (e.g. choose a /44 for the regional level prefix and allocate two /44 to the over-sized region). It should be understood that this has a negative impact on route aggregation.*

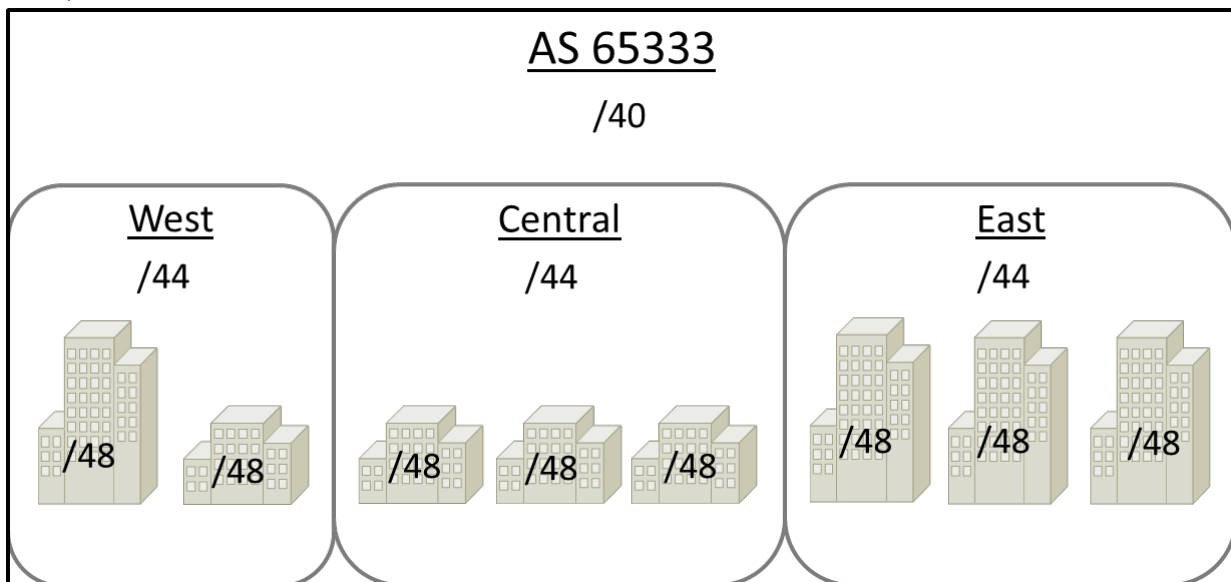


Figure 3 - Example multi-campus network

Figure 3 represents an example multi-campus enterprise network. AS 65333 has followed the general rule of allocating a /48 to each site. Since the largest campus (or region) has only three sites, they can simply round up to the next nibble boundary for their campus allocations. Then again at the AS level, a simple bump to the next nibble aligned prefix is needed. This scheme allows for plenty of growth as each campus can expand up to 15 sites and the AS can expand up to 16 (13 additional) campuses before

needing to add additional resources. Remember that within each campus, one /48 is reserved for infrastructure.

Figure 4 illustrates an example of a residential internet Service Provider (ISP) network. In this example, the operator has chosen to allocate a /48 prefix for each Multiple Dwelling Unit (MDU) but to allocate a smaller /56 prefix for individual Single Family Homes (SFH).

*This is an example for demonstrative purposes only. Individual operators will need to determine their own prefix size preference for serving customers (internal or external). The SMEs of this BCOP highly recommend a /48 for any site that requires more than one subnet and that a site be defined as an individual customer in residential networks.*

Each of the ISPs Points of Presence (PoP) serves a different number of MDUs and SFHs. For example, the area served by PoP 3 contains 50 MDUs and 150,000 SFHs. Adding up the direct need we can see that to allocate a /48 to each of the 50 MDUs, this provider needs at least a /42. To allocate a /56 to each of the 150,000 SFHs, they need at minimum a /38. This aggregate demand, rounded up to a nibble boundary, determines the need for a /36 to serve this PoP (don't forget to add the /48 for infrastructure). Since none of the other 3 PoPs require more than a /36, that was chosen as the standard PoP prefix length. Since there are only 4 PoPs in this ISPs network, they set their AS allocation at the next nibble boundary; a /32.

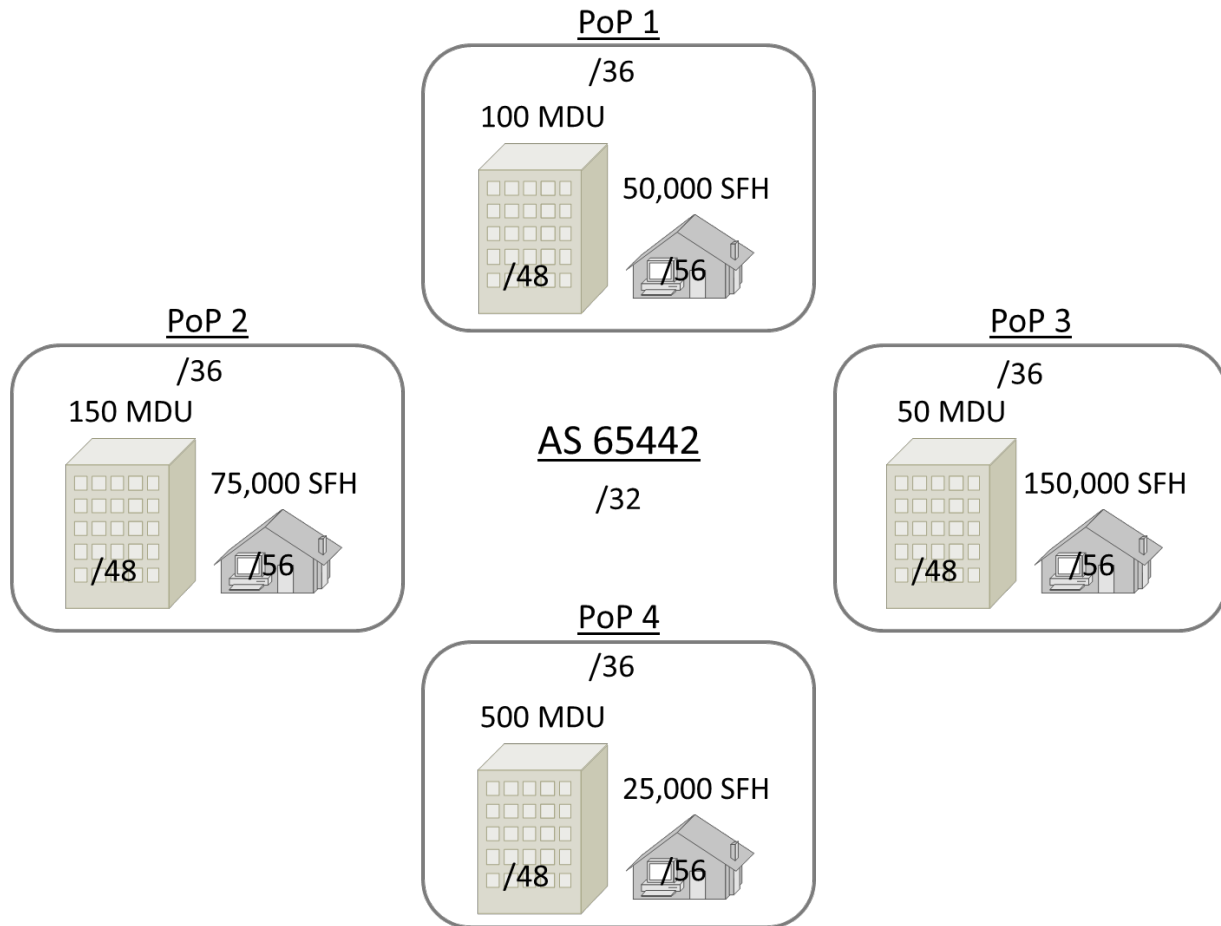


Figure 4 - Example residential ISP network

This same logic holds true no matter neither the size of your network nor the levels of hierarchy. You should always determine aggregate need of your largest site or region (including a /48 for infrastructure), round that up to a nibble-aligned prefix, and allocate an equal prefix to all sites/regions/etc. Then you repeat the exercise for each level of hierarchy within your network.

**a. Each “site” should likewise have an equalized internal hierarchy**

Within a single site, a similar hierarchy should be implemented. In this case, each individual network segment requires a /64 and departments, rooms, floors, etc. can occupy the next level up with a /60, /56, or /52 depending on whether depth or breadth better fits the network’s needs.

#### 4. BCOP Conclusion

1. Every individual network segment requires at a minimum, one /64 prefix
2. Only subnet on nibble boundaries
3. Implement a hierarchical addressing plan to allow for aggregation
  - a. Each individual site should be allocated a /48 prefix
4. One /48 from each region should be reserved for infrastructure
  - a. Loopbacks should be allocated from the top /64
  - b. Point-to-point links should be allocated a /64 and configured with a /126 or /127
5. Sites/PoPs/locations and regions, etc. should be laid out such that within each level of the hierarchy, each subnet prefix is of equal size
  - a. Each “site” should likewise have an equalized internal hierarchy