Network Working Group
Request for Comments: 4632
BCP: 122
Obsoletes: 1519
Category: Best Current Practice

V. Fuller
Cisco Systems
T. Li
Tropos Networks

August 2006

Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan

Status of This Memo

This document specifies an Internet Best Current Practices for the Internet Community, and requests discussion and suggestions for improvements. Distribution of this memo is unlimited.

Copyright Notice

Abstract

This memo discusses the strategy for address assignment of the existing 32-bit IPv4 address space with a view toward conserving the address space and limiting the growth rate of global routing state. This document obsoletes the original Classless Inter-domain Routing (CIDR) spec in RFC 1519, with changes made both to clarify the concepts it introduced and, after more than twelve years, to update the Internet community on the results of deploying the technology described.

Table of Contents

1. Introduction

This memo discusses the strategy for address assignment of the existing 32-bit IPv4 address space with a view toward conserving the address space and limiting the growth rate of global routing state. This document obsoletes the original CIDR spec [RFC1519], with changes made both to clarify the concepts it introduced and, after more than twelve years, to update the Internet community on the results of deploying the technology described.

2. History and Problem Description

What is now known as the Internet started as a research project in the 1970s to design and develop a set of protocols that could be used with many different network technologies to provide a seamless, end-to-end facility for interconnecting a diverse set of end systems. When it was determined how the 32-bit address space would be used, certain assumptions were made about the number of organizations to be connected, the number of end systems per organization, and total number of end systems on the network. The end result was the establishment (see [RFC791]) of three classes of networks: Class A (most significant address bits '00'), with 128 possible networks each and 16777216 end systems (minus special bit values reserved for network/broadcast addresses); Class B (MSB '10'), with 16384 possible networks each with 65536 end systems (less reserved values); and Class C (MSB '110'), and 2097152 possible networks each and 254 end systems (256 bit combinations minus the reserved all-zeros and all-ones patterns). The set of addresses with MSB '111' was reserved for future use; parts of this were eventually defined (MSB '1110') for use with IPv4 multicast and parts are still reserved as of the writing of this document.

In the late 1980s, the expansion and commercialization of the former research network resulted in the connection of many new organizations to the rapidly growing Internet, and each new organization required an address assignment according to the Class A/B/C addressing plan. As demand for new network numbers (particularly in the Class B space) took what appeared to be an exponential growth rate, some members of the operations and engineering community started to have concerns over the long-term scaling properties of the class A/B/C system and began thinking about how to modify network number assignment policy and routing protocols to accommodate the growth. In November, 1991, the Internet Engineering Task Force (IETF) created the ROAD (Routing and Addressing) group to examine the situation. This group met in January 1992 and identified three major problems:

1. Exhaustion of the Class B network address space. One fundamental cause of this problem is the lack of a network class of a size that is appropriate for mid-sized organization. Class C, with a maximum of 254 host addresses, is too small, whereas Class B, which allows up to 65534 host addresses, is too large for most organizations but was the best fit available for use with subnetting.

2. Growth of routing tables in Internet routers beyond the ability of current software, hardware, and people to effectively manage.

3. Eventual exhaustion of the 32-bit IPv4 address space.

   It was clear that then-current rates of Internet growth would cause the first two problems to become critical sometime between 1993 and 1995. Work already in progress on topological assignment of addressing for Connectionless Network Service (CLNS), which was presented to the community at the Boulder IETF in December of 1990, led to thoughts on how to re-structure the 32-bit IPv4 address space to increase its lifespan. Work in the ROAD group followed and eventually resulted in the publication of [RFC1338], and later, [RFC1519].

   The design and deployment of CIDR was intended to solve these problems by providing a mechanism to slow the growth of global routing tables and to reduce the rate of consumption of IPv4 address space. It did not and does not attempt to solve the third problem, which is of a more long-term nature; instead, it endeavors to ease enough of the short- to mid-term difficulties to allow the Internet to continue to function efficiently while progress is made on a longer-term solution.

   More historical background on this effort and on the ROAD group may be found in [RFC1380] and at [LWRD].

3. Classless Addressing as a Solution

The solution that the community created was to deprecate the Class A/B/C network address assignment system in favor of using "classless", hierarchical blocks of IP addresses (referred to as prefixes). The assignment of prefixes is intended to roughly follow the underlying Internet topology so that aggregation can be used to facilitate scaling of the global routing system. One implication of this strategy is that prefix assignment and aggregation is generally done according to provider-subscriber relationships, since that is how the Internet topology is determined.

When originally proposed in [RFC1338] and [RFC1519], this addressing plan was intended to be a relatively short-term response, lasting approximately three to five years, during which a more permanent addressing and routing architecture would be designed and implemented. As can be inferred from the dates on the original documents, CIDR has far outlasted its anticipated lifespan and has become the mid-term solution to the problems described above.

Note that in the following text we describe the current policies and procedures that have been put in place to implement the allocation architecture discussed here. This description is not intended to be interpreted as direction to IANA.

Coupled with address management strategies implemented by the Regional Internet Registries (see [NRO] for details), the deployment of CIDR-style addressing has also reduced the rate at which IPv4 address space has been consumed, thus providing short- to medium-term relief to problem #3, described above.

Note that, as defined, this plan neither requires nor assumes the re-assignment of those parts of the legacy "Class C" space that are not amenable to aggregation (sometimes called "the swamp"). Doing so would somewhat reduce routing table sizes (current estimate is that "the swamp" contains approximately 15,000 entries), though at a significant renumbering cost. Similarly, there is no hard requirement that any end site renumber when changing transit service provider, but end sites are encouraged do so to eliminate the need for explicit advertisement of their prefixes into the global routing system.

3.1. Basic Concept and Prefix Notation

In the simplest sense, the change from Class A/B/C network numbers to classless prefixes is to make explicit which bits in a 32-bit IPv4 address are interpreted as the network number (or prefix) associated with a site and which are the used to number individual end systems within the site. In CIDR notation, a prefix is shown as a 4-octet quantity, just like a traditional IPv4 address or network number, followed by the "/" (slash) character, followed by a decimal value between 0 and 32 that describes the number of significant bits.

For example, the legacy "Class B" network 172.16.0.0, with an implied network mask of 255.255.0.0, is defined as the prefix 172.16.0.0/16, the "/16" indicating that the mask to extract the network portion of the prefix is a 32-bit value where the most significant 16 bits are ones and the least significant 16 bits are zeros. Similarly, the legacy "Class C" network number 192.168.99.0 is defined as the prefix 192.168.99.0/24; the most significant 24 bits are ones and the least significant 8 bits are zeros.

Using classless prefixes with explicit prefix lengths allows much more flexible matching of address space blocks according to actual need. Where formerly only three network sizes were available, prefixes may be defined to describe any power of two-sized block of between one and 2^32 end system addresses. In practice, the unallocated pool of addresses is administered by the Internet Assigned Numbers Authority ([IANA]). The IANA makes allocations from this pool to Regional Internet Registries, as required. These allocations are made in contiguous bit-aligned blocks of 2^24 addresses (a.k.a. /8 prefixes). The Regional Internet Registries (RIRs), in turn, allocate or assign smaller address blocks to Local Internet Registries (LIRs) or Internet Service Providers (ISPs). These entities may make direct use of the assignment (as would commonly be the case for an ISP) or may make further sub-allocations of addresses to their customers. These RIR address assignments vary according to the needs of each ISP or LIR. For example, a large ISP might be allocated an address block of 2^17 addresses (a /15 prefix), whereas a smaller ISP may be allocated an address block of 2^11 addresses (a /21 prefix).

Note that the terms "allocate" and "assign" have specific meaning in the Internet address registry system; "allocate" refers to the delegation of a block of address space to an organization that is expected to perform further sub-delegations, and "assign" is used for sites that directly use (i.e., number individual hosts) the block of addresses received.

The following table provides a convenient shortcut to all the CIDR prefix sizes, showing the number of addresses possible in each prefix and the number of prefixes of that size that may be numbered in the 32-bit IPv4 address space:

notation	addrs/block	# blocks	
n.n.n.n/32	1	4294967296	"host route"
n.n.n.x/31	2	2147483648	"p2p link"
n.n.n.x/30	4	1073741824	
n.n.n.x/29	8	536870912	
n.n.n.x/28	16	268435456	
n.n.n.x/27	32	134217728	
n.n.n.x/26	64	67108864	
n.n.n.x/25	128	33554432	
n.n.n.0/24	256	16777216	legacy "Class C"
n.n.x.0/23	512	8388608	
n.n.x.0/22	1024	4194304	
n.n.x.0/21	2048	2097152	
n.n.x.0/20	4096	1048576	
n.n.x.0/19	8192	524288	
n.n.x.0/18	16384	262144	
n.n.x.0/17	32768	131072	
n.n.0.0/16	65536	65536	legacy "Class B"
n.x.0.0/15	131072	32768	
n.x.0.0/14	262144	16384	
n.x.0.0/13	524288	8192	
n.x.0.0/12	1048576	4096	
n.x.0.0/11	2097152	2048	
n.x.0.0/10	4194304	1024	
n.x.0.0/9	8388608	512	
n.0.0.0/8	16777216	256	legacy "Class A"
x.0.0.0/7	33554432	128	
x.0.0.0/6	67108864	64	
x.0.0.0/5	134217728	32	
x.0.0.0/4	268435456	16	
x.0.0.0/3	536870912	8	
x.0.0.0/2	1073741824	4	
x.0.0.0/1	2147483648	2	
0.0.0.0/0	4294967296	1	"default route"

n is an 8-bit decimal octet value. Point-to-point links are discussed in more detail in [RFC3021].

x is a 1- to 7-bit value, based on the prefix length, shifted into the most significant bits of the octet and converted into decimal form; the least significant bits of the octet are zero.

In practice, prefixes of length shorter than 8 have not been allocated or assigned to date, although routes to such short prefixes may exist in routing tables if or when aggressive aggregation is performed. As of the writing of this document, no such routes are seen in the global routing system, but operator error and other events have caused some of them (i.e., 128.0.0.0/1 and 192.0.0.0/2) to be observed in some networks at some times in the past.

4. Address Assignment and Routing Aggregation

Classless addressing and routing was initially developed primarily to improve the scaling properties of routing on the global Internet. Because the scaling of routing is very tightly coupled to the way that addresses are used, deployment of CIDR had implications for the way in which addresses were assigned.

4.1. Aggregation Efficiency and Limitations

The only commonly understood method for reducing routing state on a packet-switched network is through aggregation of information. For CIDR to succeed in reducing the size and growth rate of the global routing system, the IPv4 address assignment process needed to be changed to make possible the aggregation of routing information along topological lines. Since, in general, the topology of the network is determined by the service providers who have built it, topologically significant address assignments are necessarily service-provider oriented.

Aggregation is simple for an end site that is connected to one service provider: it uses address space assigned by its service provider, and that address space is a small piece of a larger block allocated to the service provider. No explicit route is needed for the end site; the service provider advertises a single aggregate route for the larger block. This advertisement provides reachability and routeability for all the customers numbered in the block.

There are two, more complex, situations that reduce the effectiveness of aggregation:

- An organization that is multi-homed. Because a multi-homed organization must be advertised into the system by each of its service providers, it is often not feasible to aggregate its routing information into the address space of any one of those providers. Note that the organization still may receive its address assignment out of a service provider’s address space (which has other advantages), but that a route to the organization’s prefix is, in the most general case, explicitly advertised by all of its service providers. For this reason, the

global routing cost for a multi-homed organization is generally the same as it was prior to the adoption of CIDR. A more detailed consideration of multi-homing practices can be found in [RFC4116].

- An organization that changes service provider but does not renumber. This has the effect of "punching a hole" in one of the original service provider's aggregated route advertisements. CIDR handles this situation by requiring that the newer service provider to advertise a specific advertisement for the re-homed organization; this advertisement is preferred over provider aggregates because it is a longer match. To maintain efficiency of aggregation, it is recommended that an organization that changes service providers plan eventually to migrate its network into a an prefix assigned from its new provider's address space. To this end, it is recommended that mechanisms to facilitate such migration, such as dynamic host address assignment that uses [RFC2131]), be deployed wherever possible, and that additional protocol work be done to develop improved technology for renumbering.

Note that some aggregation efficiency gain can still be had for multi-homed sites (and, in general, for any site composed of multiple, logical IPv4 networks); by allocating a contiguous power-of-two block address space to the site (as opposed to multiple, independent prefixes), the site's routing information may be aggregated into a single prefix. Also, since the routing cost associated with assigning a multi-homed site out of a service provider's address space is no greater than the old method of sequential number assignment by a central authority, it makes sense to assign all end-site address space out of blocks allocated to service providers.

It is also worthwhile to mention that since aggregation may occur at multiple levels in the system, it may still be possible to aggregate these anomalous routes at higher levels of whatever hierarchy may be present. For example, if a site is multi-homed to two relatively small providers that both obtain connectivity and address space from the same large provider, then aggregation by the large provider of routes from the smaller networks will include all routes to the multi-homed site. The feasibility of this sort of second-level aggregation depends on whether topological hierarchy exists among a site, its directly-connected providers, and other providers to which they are connected; it may be practical in some regions of the global Internet but not in others.

Note: In the discussion and examples that follow, prefix notation is used to represent routing destinations. This is used for illustration only and does not require that routing protocols use this representation in their updates.

4.2. Distributed Assignment of Address Space

In the early days of the Internet, IPv4 address space assignment was performed by the central Network Information Center (NIC). Class A/B/C network numbers were assigned in essentially arbitrary order, roughly according to the size of the organizations that requested them. All assignments were recorded centrally, and no attempt was made to assign network numbers in a manner that would allow routing aggregation.

When CIDR was originally deployed, the central assignment authority continued to exist but changed its procedures to assign large blocks of "Class C" network numbers to each service provider. Each service provider, in turn, assigned bitmask-oriented subsets of the provider's address space to each customer. This worked reasonably well, as long as the number of service providers was relatively small and relatively constant, but it did not scale well, as the number of service providers grew at a rapid rate.

As the Internet started to expand rapidly in the 1990s, it became clear that a single, centralized address assignment authority was problematic. This function began being de-centralized when address space assignment for European Internet sites was delegated in bit-aligned blocks of 16777216 addresses (what CIDR would later define as a /8) to the RIPE NCC ([RIPE]), effectively making it the first of the RIRs. Since then, address assignment has been formally distributed as a hierarchical function with IANA, the RIRs, and the service providers. Removing the bottleneck of a single organization having responsibility for the global Internet address space greatly improved the efficiency and response time for new assignments.

Hierarchical delegation of addresses in this manner implies that sites with addresses assigned out of a given service provider are, for routing purposes, part of that service provider and will be routed via its infrastructure. This implies that routing information about multi-homed organizations (i.e., organizations connected to more than one network service provider) will still need to be known by higher levels in the hierarchy.

A historical perspective on these issues is described in [RFC1518]. Additional discussion may also be found in [RFC3221].

5. Routing Implementation Considerations

With the change from classful network numbers to classless prefixes, it is not possible to infer the network mask from the initial bit pattern of an IPv4 address. This has implications for how routing information is stored and propagated. Network masks or prefix lengths must be explicitly carried in routing protocols. Interior routing protocols, such as OSPF [RFC2328], Intermediate System to Intermediate System (IS-IS) [RFC1195], RIPv2 [RFC2453], and Cisco Enhanced Interior Gateway Routing Protocol (EIGRP), and the BGP4 exterior routing protocol [RFC4271], all support this functionality, having been developed or modified as part of the deployment of classless inter-domain routing during the 1990s.

Older interior routing protocols, such as RIP [RFC1058], HELLO, and Cisco Interior Gateway Routing Protocol (IGRP), and older exterior routing protocols, such as Exterior Gateway Protocol (EGP) [RFC904], do not support explicit carriage of prefix length/mask and thus cannot be effectively used on the Internet other than in very limited stub configurations. Although their use may be appropriate in simple legacy end-site configurations, they are considered obsolete and should NOT be used in transit networks connected to the global Internet.

Similarly, routing and forwarding tables in layer-3 network equipment must be organized to store both prefix and prefix length or mask. Equipment that organizes its routing/forwarding information according to legacy Class A/B/C network/subnet conventions cannot be expected to work correctly on networks connected to the global Internet; use of such equipment is not recommended. Fortunately, very little such equipment is in use today.

5.1. Rules for Route Advertisement

1. Forwarding in the Internet is done on a longest-match basis. This implies that destinations that are multi-homed relative to a routing domain must always be explicitly announced into that routing domain (i.e., they cannot be summarized). If a network is multi-homed, all of its paths into a routing domain that is "higher" in the hierarchy of networks must be known to the "higher" network).

2. A router that generates an aggregate route for multiple, more-specific routes must discard packets that match the aggregate route, but not any of the more-specific routes. In other words, the "next hop" for the aggregate route should be the null destination. This is necessary to prevent forwarding loops when some addresses covered by the aggregate are not reachable.

Note that during failures, partial routing of traffic to a site that takes its address space from one service provider but that is actually reachable only through another (i.e., the case of a site that has changed service providers) may occur because such traffic will be forwarded along the path advertised by the aggregated route. Rule #2 will prevent packet misdelivery by causing such traffic to be discarded by the advertiser of the aggregated route, but the output of "traceroute" and other similar tools will suggest that a problem exists within that network rather than in the network that is no longer advertising the more-specific prefix. This may be confusing to those trying to diagnose connectivity problems; see the example in Section 6.2 for details. A solution to this perceived "problem" is beyond the scope of this document; it lies with better education of the user/operator community, not in routing technology.

An implementation following these rules should also be generalized, so that an arbitrary network number and mask are accepted for all routing destinations. The only outstanding constraint is that the mask must be left contiguous. Note that the degenerate route to prefix 0.0.0.0/0 is used as a default route and MUST be accepted by all implementations. Further, to protect against accidental advertisements of this route via the inter-domain protocol, this route should only be advertised to another routing domain when a router is explicitly configured to do so, never as a non-configured, "default" option.

5.2. How the Rules Work

Rule #1 guarantees that the forwarding algorithm used is consistent across routing protocols and implementations. Multi-homed networks are always explicitly advertised by every service provider through which they are routed, even if they are a specific subset of one service provider's aggregate (if they are not, they clearly must be explicitly advertised). It may seem as if the "primary" service provider could advertise the multi-homed site implicitly as part of its aggregate, but longest-match forwarding causes this not to work. More details are provided in [RFC4116].

Rule #2 guarantees that no routing loops form due to aggregation. Consider a site that has been assigned 192.168.64/19 by its "parent" provider, which has 192.168.0.0/16. The "parent" network will advertise 192.168.0.0/16 to the "child" network. If the "child" network were to lose internal connectivity to 192.168.65.0/24 (which is part of its aggregate), traffic from the "parent" to the to the "child" destined for 192.168.65.1 will follow the "child's" advertised route. When that traffic gets to the "child", however, the child *must not* follow the route 192.168.0.0/16 back up to the "parent", since that would result in a forwarding loop. Rule #2 says

that the "child" may not follow a less-specific route for a destination that matches one of its own aggregated routes (typically, this is implemented by installing a "discard" or "null" route for all aggregated prefixes that one network advertises to another). Note that handling of the "default" route (0.0.0.0/0) is a special case of this rule; a network must not follow the default to destinations that are part of one of its aggregated advertisements.

5.3. A Note on Prefix Filter Formats

Systems that process route announcements must be able to verify that information that they receive is acceptable according to policy rules. Implementations that filter route advertisements must allow masks or prefix lengths in filter elements. Thus, filter elements that formerly were specified as

```
accept 172.16.0.0
accept 172.25.120.0.0
accept 172.31.0.0
deny 10.2.0.0
accept 10.0.0.0
```

now look something like this:

```
accept 172.16.0.0/16
accept 172.25.0.0/16
accept 172.31.0.0/16
deny 10.2.0.0/16
accept 10.0.0.0/8
```

This is merely making explicit the network mask that was implied by the Class A/B/C classification of network numbers. It is also useful to enhance filtering capability to allow the match of a prefix and all more-specific prefixes with the same bit pattern; fortunately, this functionality has been implemented by most vendors of equipment used on the Internet.

5.4. Responsibility for and Configuration of Aggregation

Under normal circumstances, a routing domain (or "Autonomous System") that has been allocated or assigned a set of prefixes has sole responsibility for aggregation of those prefixes. In the usual case, the AS will install configuration in one or more of its routers to generate aggregate routes based on more-specific routes known to its internal routing system. These aggregate routes are advertised into the global routing system by the border routers for the routing domain. The more-specific internal routes that overlap with the aggregate routes should not be advertised globally. In some cases,

an AS may wish to delegate aggregation responsibility to another AS (for example, a customer may wish for its service provider to generate aggregated routing information on its behalf); in such cases, aggregation is performed by a router in the second AS according to the routes that it receives from the first, combined with configured policy information describing how those routes should be aggregated.

Note that one provider may choose to perform aggregation on the routes it receives from another without explicit agreement; this is termed "proxy aggregation". This can be a useful tool for reducing the amount of routing state that an AS must carry and propagate to its customers and neighbors. However, proxy aggregation can also create unintended consequences in traffic engineering. Consider what happens if both AS 2 and 3 receive routes from AS 1 but AS 2 performs proxy aggregation while AS 3 does not. Other ASes that receive transit routing information from both AS 2 and AS 3 will see an inconsistent view of the routing information originated by AS 1. This may cause an unexpected shift of traffic toward AS 1 through AS 3 for AS 3's customers and any others receiving transit routes from AS 3. Because proxy aggregation can cause unanticipated consequences for parts of the Internet that have no relationship with either the source of the aggregated routes or the party providing aggregation, it should be used with extreme caution.

Configuration of the routes to be combined into aggregates is an implementation of routing policy and requires some manually maintained information. As an addition to the information that must be maintained for a set of routeable prefixes, aggregation configuration is typically just a line or two defining the range of the block of IPv4 addresses to be aggregated. A site performing its own aggregation is doing so for address blocks that it has been assigned; a site performing aggregation on behalf of another knows this information because of an agreement to delegate aggregation. Assuming that the best common practice for network administrators is to exchange lists of prefixes to accept from each other, configuration of aggregation information does not introduce significant additional administrative overhead.

The generation of an aggregate route is usually specified either statically or in response to learning an active dynamic route for a prefix contained within the aggregate route. If such dynamic aggregate route advertisement is done, care should be taken that routes are not excessively added or withdrawn (known as "route flapping"). In general, a dynamic aggregate route advertisement is added when at least one component of the aggregate becomes reachable and it is withdrawn only when all components become unreachable. Properly configured, aggregated routes are more stable than non-aggregated routes and thus improve global routing stability.

Implementation note: Aggregation of the "Class D" (multicast) address space is beyond the scope of this document.

5.5. Route Propagation and Routing Protocol Considerations

Prior to the original deployment of CIDR, common practice was to propagate routes learned via exterior routing protocols (i.e., EGP or BGP) through a site's interior routing protocol (typically, OSPF, IS-IS, or RIP). This was done to ensure that consistent and correct exit points were chosen for traffic to be sent to a destination learned through those protocols. Four evolutionary effects -- the advent of CIDR, explosive growth of global routing state, widespread adoption of BGP4, and a requirement to propagate full path information -- have combined to deprecate that practice. To ensure proper path propagation and prevent inter-AS routing inconsistency (BGP4's loop detection/prevention mechanism requires full path propagation), transit networks must use internal BGP (iBGP) for carrying routes learned from other providers both within and through their networks.

6. Example of New Address Assignments and Routing

6.1. Address Delegation

Consider the block of 524288 (2^19) addresses, beginning with 10.24.0.0 and ending with 10.31.255.255, allocated to a single network provider, "PA". This is equivalent in size to a block of 2048 legacy "Class C" network numbers (or /24s). A classless route to this block would be described as 10.24.0.0 with a mask of 255.248.0.0 and the prefix 10.24.0.0/13.

Assume that this service provider connects six sites in the following order (significant because it demonstrates how temporary "holes" may form in the service provider's address space):

- "C1", requiring fewer than 2048 addresses (/21 or 8 x /24)
- "C2", requiring fewer than 4096 addresses (/20 or 16 x /24)
- "C3", requiring fewer than 1024 addresses (/22 or 4 x /24)
- "C4", requiring fewer than 1024 addresses (/22 or 4 x /24)
- "C5", requiring fewer than 512 addresses (/23 or 2 x /24)
- "C6", requiring fewer than 512 addresses (/23 or 2 x /24)

In all cases, the number of IPv4 addresses "required" by each site is assumed to allow for significant growth. The service provider delegates its address space as follows:

- C1. assign 10.24.0 through 10.24.7. This block of networks is described by the route 10.24.0.0/21 (mask 255.255.248.0).
- C2. Assign 10.24.16 through 10.24.31. This block is described by the route 10.24.16.0/20 (mask 255.255.240.0).
- C3. Assign 10.24.8 through 10.24.11. This block is described by the route 10.24.8.0/22 (mask 255.255.252.0).
- C4. Assign 10.24.12 through 10.24.15. This block is described by the route 10.24.12.0/22 (mask 255.255.252.0).
- C5. Assign 10.24.32 and 10.24.33. This block is described by the route 10.24.32.0/23 (mask 255.255.254.0).
- C6. Assign 10.24.34 and 10.24.35. This block is described by the route 10.24.34.0/23 (mask 255.255.254.0).

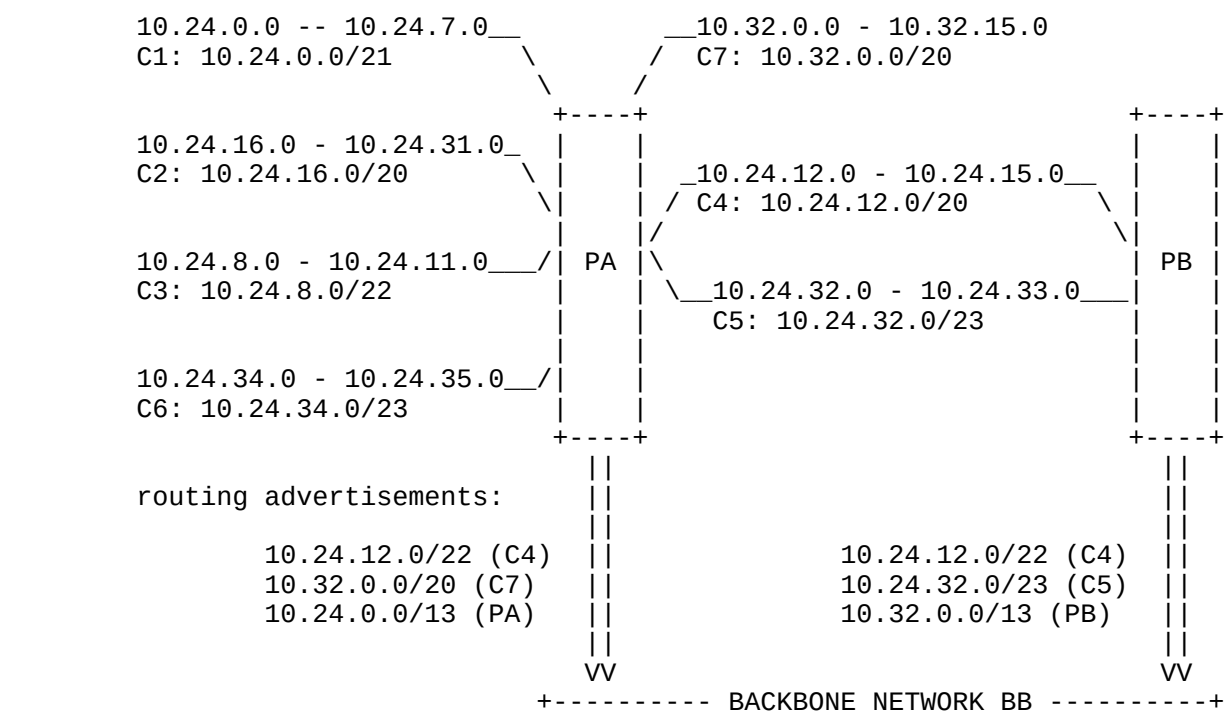
These six sites should be represented as six prefixes of varying size within the provider’s IGP. If, for some reason, the provider uses an obsolete IGP that doesn’t support classless routing or variable-length subnets, then explicit routes for all /24s will have to be carried.

To make this example more realistic, assume that C4 and C5 are multi-homed through some other service provider, "PB". Further assume the existence of a site, "C7", that was originally connected to "RB" but that has moved to "PA". For this reason, it has a block of network numbers that are assigned out PB’s block of (the next) 2048 x /24.

- C7. Assign 10.32.0 through 10.32.15. This block is described by the route 10.32.0.0/20 (mask 255.255.240.0).

For the multi-homed sites, assume that C4 is advertised as primary via "RA" and secondary via "RB"; and that C5 is primary via "RB" and secondary via "RA". In addition, assume that "RA" and "RB" are both connected to the same transit service provider, "BB".

Graphically, this topology looks something like this:



6.2. Routing Advertisements

To follow rule #1, PA will need to advertise the block of addresses that it was given and C7. Since C4 is multi-homed and primary through PA, it must also be advertised. C5 is multi-homed and primary through PB. In principle (and in the example above), it need not be advertised, since longest match by PB will automatically select PB as primary and the advertisement of PA's aggregate will be used as a secondary. In actual practice, C5 will normally be advertised via both providers.

Advertisements from "PA" to "BB" will be

```
10.24.12.0/22 primary    (advertises C4)
10.32.0.0/20 primary     (advertises C7)
10.24.0.0/13 primary     (advertises remainder of PA)
```

For PB, the advertisements must also include C4 and C5, as well as its block of addresses.

Advertisements from "PB" to "BB" will be

```
10.24.12.0/22 secondary  (advertises C4)
10.24.32.0/23 primary    (advertises C5)
10.32.0.0/13 primary     (advertises remainder of RB)
```

To illustrate the problem diagnosis issue mentioned in Section 5.1, consider what happens if PA loses connectivity to C7 (the site that is assigned out of PB's space). In a stateful protocol, PA will announce to BB that 10.32.0.0/20 has become unreachable. Now, when BB flushes this information out of its routing table, any future traffic sent through it for this destination will be forwarded to PB (where it will be dropped according to Rule #2) by virtue of PB's less-specific match, 10.32.0.0/13. Although this does not cause an operational problem (C7 is unreachable in any case), it does create some extra traffic across "BB" (and may also prove confusing to someone trying to debug the outage with "traceroute"). A mechanism to cache such unreachable state might be nice, but it is beyond the scope of this document.

7. Domain Name Service Considerations

One aspect of Internet services that was notably affected by the move to CIDR was the mechanism used for address-to-name translation: the IN-ADDR.ARPA zone of the domain system. Because this zone is delegated on octet boundaries only, the move to an address assignment plan that uses bitmask-oriented addressing caused some increase in work for those who maintain parts of the IN-ADDR.ARPA zone.

A description of techniques to populate the IN-ADDR.ARPA zone when and used address that blocks that do not align to octet boundaries is described in [RFC2317].

8. Transition to a Long-Term Solution

CIDR was designed to be a short-term solution to the problems of routing state and address depletion on the IPv4 Internet. It does not change the fundamental Internet routing or addressing architectures. It is not expected to affect any plans for transition to a more long-term solution except, perhaps, by delaying the urgency of developing such a solution.

9. Analysis of CIDR's Effect on Global Routing State

When CIDR was first proposed in the early 1990s, the original authors made some observations about the growth rate of global routing state and offered projections on how CIDR deployment would, hopefully, reduce what appeared to be exponential growth to a more sustainable rate. Since that deployment, an ongoing effort, called "The CIDR Report" [CRPT], has attempted to quantify and track that growth rate. What follows is a brief summary of the CIDR report as of March 2005, with an attempt to explain the various patterns and changes of growth rate that have occurred since measurements of the size of global routing state began in 1988.

When the graph of "Active BGP Table Entries" [CBGP] is examined, there appear to be several different growth trends with distinct inflection points reflecting changes in policy and practice. The trends and events that are believed to have caused them were as follows:

1. Exponential growth at the far left of the graph. This represents the period of early expansion and commercialization of the former research network, from the late 1980s through approximately 1994. The major driver for this growth was a lack of aggregation capability for transit providers, and the widespread use of legacy Class C allocations for end sites. Each time a new site was connected to the global Internet, one or more new routing entries were generated.

2. Acceleration of the exponential trend in late 1993 and early 1994 as CIDR "supernet" blocks were first assigned by the NIC and routed as separate legacy class-C networks by service provider.

3. A sharp drop in 1994 as BGP4 deployment by providers allowed aggregation of the "supernet" blocks. Note that the periods of largest declines in the number of routing table entries typically correspond to the weeks following each meeting of the IETF CIDR Deployment Working Group.

4. Roughly linear growth from mid-1994 to early 1999 as CIDR-based address assignments were made and aggregated routes added throughout the network.

5. A new period of exponential growth again from early 1999 until 2001 as the "high-tech bubble" fueled both rapid expansion of the Internet, as well as a large increase in more-specific route advertisements for multi-homing and traffic engineering.

6. Flattening of growth through 2001 caused by a combination of the "dot-com bust", which caused many organizations to cease operations, and the "CIDR police" [CPOL] work aimed at improving aggregation efficiency.

7. Roughly linear growth through 2002 and 2003. This most likely represents a resumption of the "normal" growth rate observed before the "bubble", as well as an end to the "CIDR Police" effort.

8. A more recent trend of exponential growth beginning in 2004. The best explanation would seem to be an improvement of the global economy driving increased expansion of the Internet and the continued absence of the "CIDR Police" effort, which previously served as an educational tool for new providers to improve aggregation efficiency. There have also been some cases where service providers have deliberately de-aggregated prefixes in an attempt to mitigate security problems caused by conflicting route advertisements (see Section 12). Although this behavior may solve the short-term problems seen by such providers, it is fundamentally non-scalable and quite detrimental to the community as a whole. In addition, there appear to be many providers advertising both their allocated prefixes and all the /24 components thereof, probably due to a lack of consistent current information about recommended routing configuration.

10. Conclusions and Recommendations

In 1992, when CIDR was first developed, there were serious problems facing the continued growth of the Internet. Growth in routing state complexity and the rapid increase in consumption of address space made it appear that one or both problems would preclude continued growth of the Internet within a few short years.

Deployment of CIDR, in combination with BGP4's support for carrying classless prefix routes, alleviated the short-term crisis. It was only through a concerted effort by both the equipment manufacturers and the provider community that this was achieved. The threat (and, perhaps in some cases, actual implementation of) charging networks for advertising prefixes may have offered an additional incentive to share the address space, and thus the associated costs of advertising routes to service providers.

The IPv4 routing system architecture carries topology information based on aggregate address advertisements and a collection of more-specific advertisements that are associated with traffic engineering, multi-homing, and local configuration. As of March 2005, the base aggregate address load in the routing system has some 75,000 entries.

Approximately 85,000 additional entries are more specific entries of this base "root" collection. There is reason to believe that many of these additional entries exist to solve problems of regional or even local scope and should not need to be globally propagated.

An obvious question to ask is whether CIDR can continue to be a viable approach to keeping global routing state growth and address space depletion at sustainable rates. Recent measurements indicate that exponential growth has resumed, but further analysis suggests that this trend can be mitigated by a more active effort to educate service providers as to efficient aggregation strategies and proper equipment configuration. Looking farther forward, there is a clear need for better multi-homing technology that does not require global routing state for each site and for methods of performing traffic load balancing that do not require adding even more state. Without such developments and in the absence of major architectural change, aggregation is the only tool available for making routing scale in the global Internet.

11. Status Updates to CIDR Documents

This memo renders obsolete and requests re-classification as Historic the following RFCs describing CIDR usage and deployment:

- RFC 1467: Status of CIDR Deployment in the Internet

  This Informational RFC described the status of CIDR deployment in 1993. As of 2005, CIDR has been thoroughly deployed, so this status note only provides a historical data point.

- RFC 1481: IAB Recommendation for an Intermediate Strategy to Address the Issue of Scaling

  This very short Informational RFC described the IAB’s endorsement of the use of CIDR to address scaling issues. Because the goal of RFC 1481 has been achieved, it is now only of historical value.

- RFC 1482: Aggregation Support in the NSFNET Policy-Based Routing Database

  This Informational RFC describes plans for support of route aggregation, as specified by CIDR, on the NSFNET. Because the NSFNET has long since ceased to exist and CIDR has been ubiquitously deployed, RFC 1482 now only has historical relevance.

- RFC 1517: Applicability Statement for the Implementation of Classless Inter-Domain Routing (CIDR)

This Standards Track RFC described where CIDR was expected to be required and where it was expected to be (strongly) recommended. With the full deployment of CIDR on the Internet, situations where CIDR is not required are of only historical interest.

- RFC 1518: An Architecture for IP Address Allocation with CIDR

  This Standards Track RFC discussed routing and address aggregation considerations at some length. Some of these issues are summarized in this document in section Section 3.1. Because address assignment policies and procedures now reside mainly with the RIRs, it is not appropriate to try to document those practices in a Standards Track RFC. In addition, [RFC3221] also describes many of the same issues from point of view of the routing system.

- RFC 1520: Exchanging Routing Information Across Provider Boundaries in the CIDR Environment

  This Informational RFC described transition scenarios where CIDR was not fully supported for exchanging route information between providers. With the full deployment of CIDR on the Internet, such scenarios are no longer operationally relevant.

- RFC 1817: CIDR and Classful Routing

  This Informational RFC described the implications of CIDR deployment in 1995; it notes that formerly-classful addresses were to be allocated using CIDR mechanisms and describes the use of a default route for non-CIDR-aware sites. With the full deployment of CIDR on the Internet, such scenarios are no longer operationally relevant.

- RFC 1878: Variable Length Subnet Table For IPv4

  This Informational RFC provided a table of pre-calculated subnet masks and address counts for each subnet size. With the incorporation of a similar table into this document (see Section 3.1), it is no longer necessary to document it in a separate RFC.

- RFC 2036: Observations on the use of Components of the Class A Address Space within the Internet

  This Informational RFC described several operational issues associated with the allocation of classless prefixes from previously-classful address space. With the full deployment of CIDR on the Internet and more than half a dozen years of experience making classless prefix allocations out of historical "Class A" address space, this RFC now has only historical value.

12. Security Considerations

The introduction of routing protocols that support classless prefixes and a move to a forwarding model that mandates that more-specific (longest-match) routes be preferred when they overlap with routes to less-specific prefixes introduces at least two security concerns:

1. Traffic can be hijacked by advertising a prefix for a given destination that is more specific than the aggregate that is normally advertised for that destination. For example, assume that a popular end system with the address 192.168.17.100 is connected to a service provider that advertises 192.168.16.0/20. A malicious network operator interested in intercepting traffic for this site might advertise, or at least attempt to advertise, 192.168.17.0/24 into the global routing system. Because this prefix is more specific than the "normal" prefix, traffic will be diverted away from the legitimate end system and to the network owned by the malicious operator. Prior to the advent of CIDR, it was possible to induce traffic from some parts of the network to follow a false advertisement that exactly matched a particular network number; CIDR makes this problem somewhat worse, since longest-match routing generally causes all traffic to prefer more-specific routes over less-specific routes. The remedy for the CIDR-based attack, though, is the same as for a pre-CIDR-based attack: establishment of trust relationships between providers, coupled with and strong route policy filters at provider borders. Unfortunately, the implementation of such filters is difficult in the highly de-centralized Internet. As a workaround, many providers do implement generic filters that set upper bounds, derived from RIR guidelines for the sizes of blocks that they allocate, on the lengths of prefixes that are accepted from other providers. Note that "spammers" have been observed using this sort of attack to hijack address space temporarily in order to hide the origin of the traffic ("spam" email messages) that they generate.

2. Denial-of-service attacks can be launched against many parts of the Internet infrastructure by advertising a large number of routes into the system. Such an attack is intended to cause router failures by overflowing routing and forwarding tables. A good example of a non-malicious incident that caused this sort of failure was the infamous "AS 7007" event [7007], where a router mis-configuration by an operator caused a huge number of invalid routes to be propagated through the global routing system. Again, this sort of attack is not really new with CIDR; using legacy Class A/B/C routes, it was possible to advertise a maximum of 16843008 unique network numbers into the global routing system, a number that is sufficient to cause problems for even

the most modern routing equipment made in 2005. What is different is that the moderate complexity of correctly configuring routers in the presence of CIDR tends to make accidental "attacks" of this sort more likely. Measures to prevent this sort of attack are much the same as those described above for the hijacking, with the addition that best common practice is also to configure a reasonable maximum number of prefixes that a border router will accept from its neighbors.

Note that this is not intended to be an exhaustive analysis of the sorts of attacks that CIDR makes easier; a more comprehensive analysis of security vulnerabilities in the global routing system is beyond the scope of this document.

13. Acknowledgements

The authors wish to express appreciation to the other original authors of RFC 1519 (Kannan Varadhan, Jessica Yu); to the ROAD group, with whom many of the ideas behind CIDR were inspired and developed; and to the early reviewers of this re-spun version of the document (Barry Greene, Danny McPherson, Dave Meyer, Eliot Lear, Bill Norton, Ted Seely, Philip Smith, Pekka Savola), whose comments, corrections, and suggestions were invaluable. We would especially like to thank Geoff Huston for contributions well above and beyond the call of duty.

14. References

14.1. Normative References

[RFC791] Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.

14.2. Informative References

[7007] "NANOG mailing list discussion of the "AS 7007" incident", <http://www.merit.edu/mail.archives/nanog/1997-04/msg00340.html>.

[CBGP] "Graph: Active BGP Table Entries, 1988 to Present", <http://bgp.potaroo.net/as4637/>.

[CPOL] "CIDR Police - Please Pull Over and Show Us Your BGP", <http://www.nanog.org/mtg-0302/cidr.html>.

[CRPT] "The CIDR Report", <http://www.cidr-report.org/>.

[IANA] "Internet Assigned Numbers Authority", <http://www.iana.org>.

[LWRD] "The Long and Winding Road", <http://rms46.vlsm.org/1/42.html>.

[NRO] "Number Resource Organization", <http://www.nro.net>.

[RFC904] Mills, D., "Exterior Gateway Protocol formal specification", RFC 904, April 1 1984.

[RFC1058] Hedrick, C., "Routing Information Protocol", RFC 1058, June 1988.

[RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, December 1990.

[RFC1338] Fuller, V., Li, T., Yu, J., and K. Varadhan, "Supernetting: an Address Assignment and Aggregation Strategy", RFC 1338, June 1992.

[RFC1380] Gross, P. and P. Almquist, "IESG Deliberations on Routing and Addressing", RFC 1380, November 1992.

[RFC1518] Rekhter, Y. and T. Li, "An Architecture for IP Address Allocation with CIDR", RFC 1518, September 1993.

[RFC1519] Fuller, V., Li, T., Yu, J., and K. Varadhan, "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", RFC 1519, September 1993.

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

[RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[RFC2317] Eidnes, H., de Groot, G., and P. Vixie, "Classless IN-ADDR.ARPA delegation", BCP 20, RFC 2317, March 1998.

[RFC2453] Malkin, G., "RIP Version 2", STD 56, RFC 2453, November 1998.

[RFC3021] Retana, A., White, R., Fuller, V., and D. McPherson, "Using 31-Bit Prefixes on IPv4 Point-to-Point Links", RFC 3021, December 2000.

[RFC3221] Huston, G., "Commentary on Inter-Domain Routing in the Internet", RFC 3221, December 2001.

[RFC4116] Abley, J., Lindqvist, K., Davies, E., Black, B., and V. Gill, "IPv4 Multihoming Practices and Limitations", RFC 4116, July 2005.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RIPE] "RIPE Network Coordination Centre", <http://www.ripe.net>.

Authors' Addresses

Vince Fuller
170 W. Tasman Drive
San Jose, CA 95134
USA

EMail: vaf@cisco.com

Tony Li
555 Del Rey Avenue
Sunnyvale, CA 94085

Email: tli@tropos.com

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).