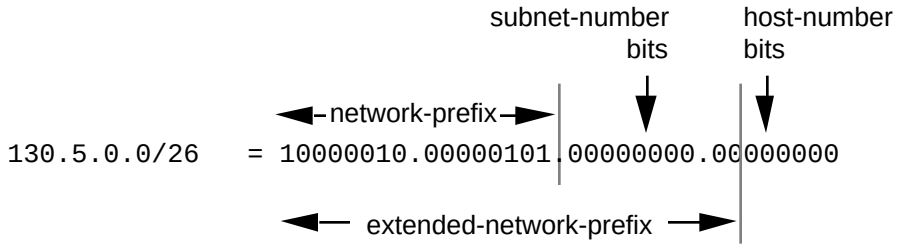

Figure 16: 130.5.0.0/16 with a /26 Extended-Network Prefix

Route Aggregation

VLSM also allows the recursive division of an organization's address space so that it can be reassembled and aggregated to reduce the amount of routing information at the top level. Conceptually, a network is first divided into subnets, some of the subnets are further divided into sub-subnets, and some of the sub-subnets are divided into sub^2-subnets. This allows the detailed structure of routing information for one subnet group to be hidden from routers in another subnet group.

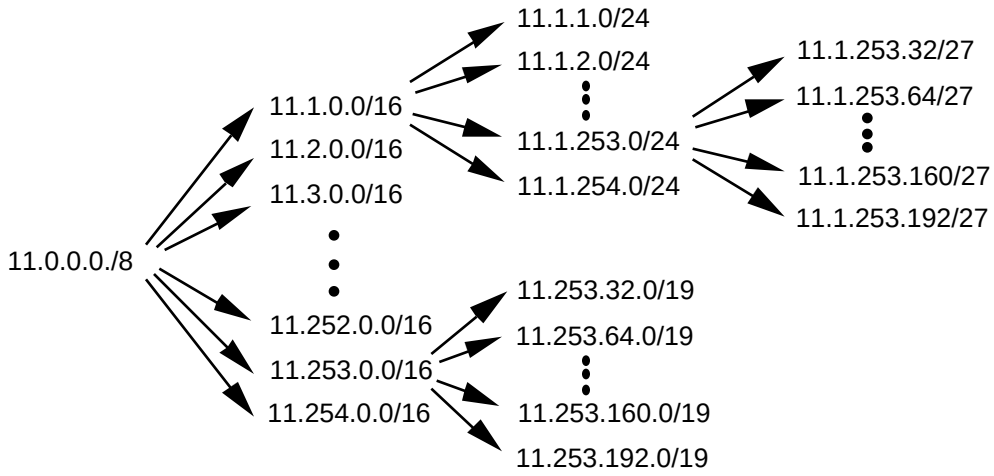

Figure 17: VLSM Permits the Recursive Division of a Network Prefix

In Figure 17, the 11.0.0.0/8 network is first configured with a /16 extended-network-prefix. The 11.1.0.0/16 subnet is then configured with a /24 extended-network-prefix and the 11.253.0.0/16 subnet is configured with a /19 extended-network-prefix. Note that the recursive process does not require that the same extended-network-prefix be assigned at each level of the recursion. Also, the recursive sub-division of the organization's address space can be carried out as far as the network administrator needs to take it.

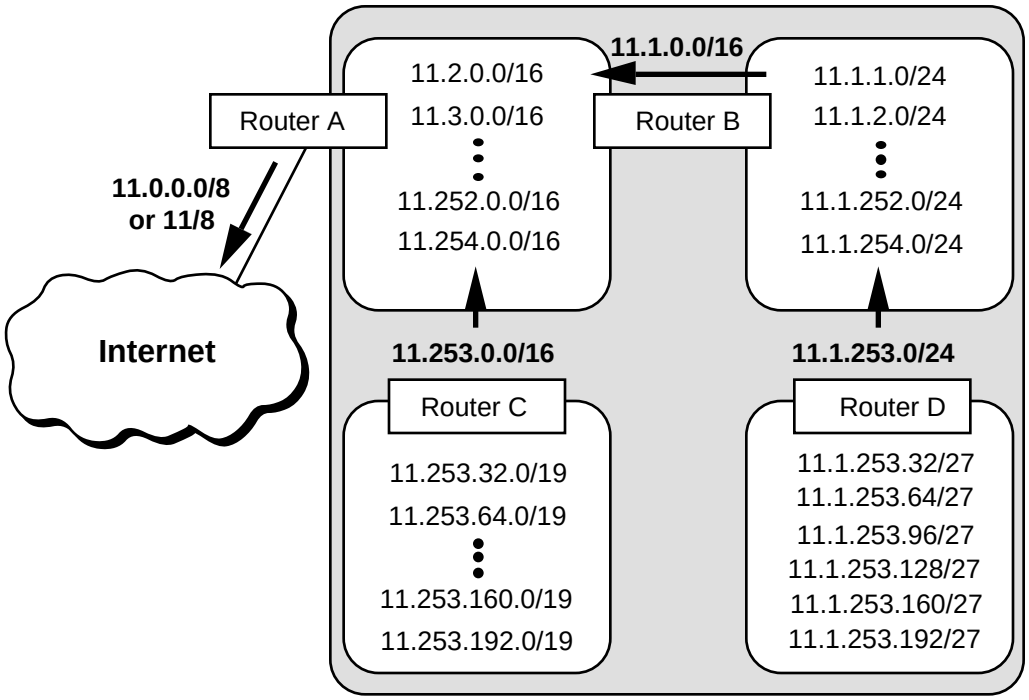

Figure 18: VLSM Permits Route Aggregation - Reducing Routing Table Size

Figure 18 illustrates how a planned and thoughtful allocation of VLSM can reduce the size of an organization's routing tables. Notice how Router D is able to summarize the six subnets behind it into a single advertisement (11.1.253.0/24) and how Router B is able to aggregate all of subnets behind it into a single advertisement. Likewise, Router C is able to summarize the six subnets behind it into a single advertisement (11.253.0.0/16). Finally, since the subnet structure is not visible outside of the organization, Router A injects a single route into the global Internet's routing table - 11.0.0.0/8 (or 11/8).

VLSM Design Considerations

When developing a VLSM design, the network designer must recursively ask the same set of questions as for a traditional subnet design. The same set of design decisions must be made at each level of the hierarchy:

1) How many total subnets does this level need today?
2) How many total subnets will this level need in the future?
3) How many hosts are there on this level's largest subnet today?
4) How many hosts will there be on this level's largest subnet be in the future?

At each level, the design team must make sure that they have enough extra bits to support the required number of sub-entities in the next and further levels of recursion.

Assume that a network is spread out over a number of sites. For example, if an organization has three campuses today it probably needs 3-bits of subnetting ($2^3 = 8$) to allow the addition of more campuses in the future. Now, within each campus, there is likely to be a secondary level of subnetting to identify each building. Finally, within each building, a third level of subnetting might identify each of the individual workgroups. Following this hierarchical model, the top level is determined by the number of campuses, the mid-level is based on the number of buildings at each site, and the lowest level is determined by the "maximum number of subnets/maximum number of users per subnet" in each building.

The deployment of a hierarchical subnetting scheme requires careful planning. It is essential that the network designers recursively work their way down through their addressing plan until they get to the bottom level. At the bottom level, they must make sure that the leaf subnets are large enough to support the required number of hosts. When the addressing plan is deployed, the addresses from each site will be aggregable into a single address block that keeps the backbone routing tables from becoming too large.

Requirements for the Deployment of VLSM

The successful deployment of VLSM has three prerequisites:

- The routing protocols must carry extended-network-prefix information with each route advertisement.
- All routers must implement a consistent forwarding algorithm based on the "longest match."
- For route aggregation to occur, addresses must be assigned so that they have topological significance.

Routing Protocols Must Carry Extended-Network-Prefix Lengths

Modern routing protocols, such as OSPF and I-IS-IS, enable the deployment of VLSM by providing the extended-network-prefix length or mask value along with each route advertisement. This permits each subnetwork to be advertised with its corresponding prefix length or mask. If the routing protocols did not carry prefix information, a router would have to either assume that the locally configured prefix length should be applied, or perform a look-up in a statically configured prefix table that contains all of the required masking information. The first alternative cannot guarantee that the correct prefix is applied, and static tables do not scale since they are difficult to maintain and subject to human error.

The bottom line is that if you want to deploy VLSM in a complex topology, you must select OSPF or I-IS-IS as the Interior Gateway Protocol (IGP) rather than RIP-1! It should be mentioned that RIP-2, defined in RFC 1388, improves the RIP protocol by allowing it to carry extended-network-prefix information. Therefore, RIP-2 supports the deployment of VLSM.

Forwarding Algorithm is Based on the "Longest Match"

All routers must implement a consistent forwarding algorithm based on the "longest match" algorithm. The deployment of VLSM means that the set of networks associated with extended-network-prefixes may manifest a subset relationship. A route with a longer extended-network-prefix describes a smaller set of destinations than the same route with a shorter extended-network-prefix. As a result, a route with a longer extended-network-prefix is said to be "more specific" while a route with a shorter extended-network-prefix is said to be "less specific." Routers must use the route with the longest matching extended-network-prefix (most specific matching route) when forwarding traffic.

For example, if a packet's destination IP address is 11.1.2.5 and there are three network prefixes in the routing table (11.1.2.0/24, 11.1.0.0/16, and 11.0.0.0/8), the router would select the route to 11.1.2.0/24. The 11.1.2.0/24 route is selected because its prefix has the greatest number of corresponding bits in the Destination IP address of the packet. This is illustrated in Figure 19.

```
   Destination    11.1.2.5    = 00001011.00000001.00000010.00000101

*  Route #1       11.1.2.0/24 = 00001011.00000001.00000010.00000000
   Route #2       11.1.0.0/16 = 00001011.00000001.00000000.00000000
   Route #3       11.0.0.0/8  = 00001011.00000000.00000000.00000000
```

Figure 19: Best Match is with the Route Having the Longest Prefix (Most Specific)

There is a very subtle but extremely important issue here. Since the destination address matches all three routes, it must be assigned to a host which is attached to the 11.1.2.0/24 subnet. If the 11.1.2.5 address is assigned to a host that is attached to the 11.1.0.0/16 or 11.0.0.0/8 subnet, the routing system will *never* route traffic to the host since the "longest match algorithm" assumes that the host is part of the 11.1.2.0/24 subnet. This means that great care must be taken when assigning host addresses to make sure that every host is reachable!

Topologically Significant Address Assignment

Since OSPF and I-IS-IS convey the extended-network-prefix information with each route, the VLSM subnets can be scattered throughout an organization's topology. However, to support hierarchical routing and reduce the size of an organization's routing tables, addresses should be assigned so that they are topologically significant.

Hierarchical routing requires that addresses be assigned to reflect the actual network topology. This reduces the amount of routing information by taking the set of addresses assigned to a particular region of the topology, and aggregating them into a single routing advertisement for the entire set. Hierarchical routing allows this to be done recursively at various points within the hierarchy of the routing topology. If addresses do not have a topological significance, aggregation cannot be performed and the size of the routing tables cannot be reduced. Remember this point when we discuss CIDR aggregation later in this paper.

VLSM Example

Given

An organization has been assigned the network number 140.25.0.0/16 and it plans to deploy VLSM. Figure 20 provides a graphic display of the VLSM design for the organization.

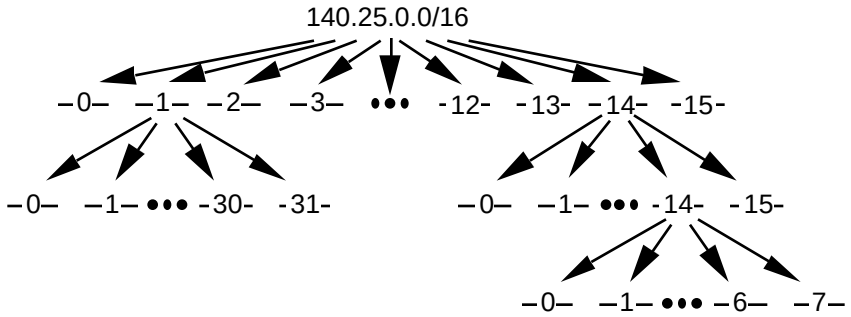

Figure 20: Address Strategy for VLSM Example

The first step of the subnetting process divides the base network address into 16 equal-sized address blocks. Then Subnet #1 is divided it into 32 equal-sized address blocks and Subnet #14 is divided into 16 equal-sized address blocks. Finally, Subnet #14-14 is divided into 8 equal-sized address blocks.

Define the 16 Subnets of 140.25.0.0/16

The first step in the subnetting process divides the base network address into 16 equal-size address blocks. This is illustrated in Figure 21.

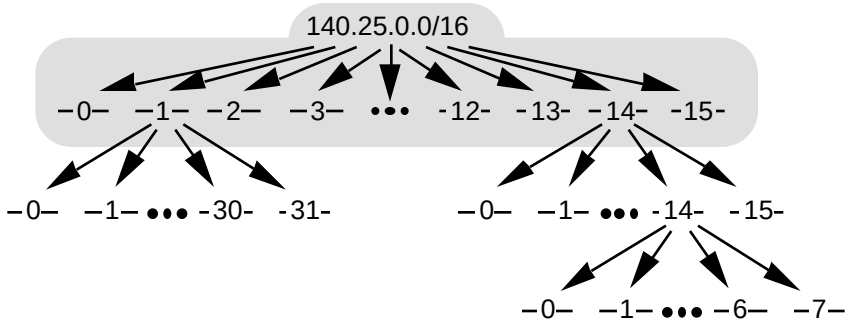

Figure 21: Define the 16 Subnets for 140.25.0.0/16

Since $16 = 2^4$, four bits are required to uniquely identify each of the 16 subnets. This means that the organization needs four more bits, or a /20, in the extended-network-prefix to define the 16 subnets of 140.25.0.0/16. Each of these subnets represents a contiguous block of 2^{12} (or 4,096) network addresses.

The 16 subnets of the 140.25.0.0/16 address block are given below. The subnets are numbered 0 through 15. The underlined portion of each address identifies the extended-network-prefix, while the **bold** digits identify the 4-bits representing the subnet-number field:

```
Base Network: 10001100.00011001.00000000.00000000 = 140.25.0.0/16

Subnet #0:    10001100.00011001.0000 0000.00000000 = 140.25.0.0/20
Subnet #1:    10001100.00011001.0001 0000.00000000 = 140.25.16.0/20
Subnet #2:    10001100.00011001.0010 0000.00000000 = 140.25.32.0/20
Subnet #3:    10001100.00011001.0011 0000.00000000 = 140.25.48.0/20
Subnet #4:    10001100.00011001.0100 0000.00000000 = 140.25.64.0/20
    :
    :
Subnet #13:   10001100.00011001.1101 0000.00000000 = 140.25.208.0/20
Subnet #14:   10001100.00011001.1110 0000.00000000 = 140.25.224.0/20
Subnet #15:   10001100.00011001.1111 0000.00000000 = 140.25.240.0/20
```

Define the Host Addresses for Subnet #3 (140.25.48.0/20)

Let's examine the host addresses that can be assigned to Subnet #3 (140.25.48.0/20). This is illustrated in Figure 22.

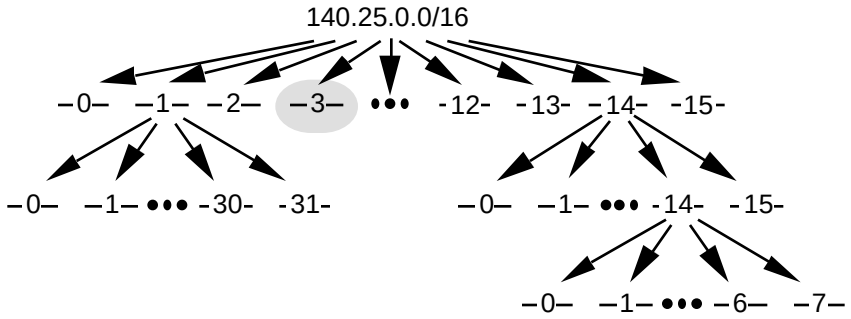

Figure 22: Define the Host Addresses for Subnet #3 (140.25.48.0/20)

Since the host-number field of Subnet #3 contains 12 bits, there are 4,094 valid host addresses (2^{12}-2) in the address block. The hosts are numbered 1 through 4,094.

The valid host addresses for Subnet #3 are given below. The underlined portion of each address identifies the extended-network-prefix, while the **bold** digits identify the 12-bit host-number field:

```
Subnet #3:    10001100.00011001.0011 0000.00000000 = 140.25.48.0/20

Host #1:    10001100.00011001.0011 0000.00000001 = 140.25.48.1/20
Host #2:    10001100.00011001.0011 0000.00000010 = 140.25.48.2/20
Host #3:    10001100.00011001.0011 0000.00000011 = 140.25.48.3/20
   :
   :
Host #4093: 10001100.00011001.0011 1111.11111101 = 140.25.63.253/20
Host #4094: 10001100.00011001.0011 1111.11111110 = 140.25.63.254/20
```

The broadcast address for Subnet #3 is the all 1's host address or:

10001100.00011001.0011**1111.11111111** = 140.25.63.255

The broadcast address for Subnet #3 is exactly one less than the base address for Subnet #4 (140.25.64.0).

Define the Sub-Subnets for Subnet #14 (140.25.224.0/20)

After the base network address is divided into sixteen subnets, Subnet #14 is further subdivided into 16 equal-size address blocks. This is illustrated in Figure 23.

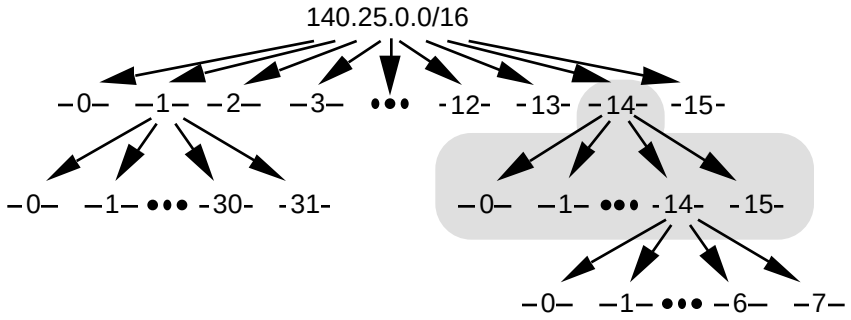

Figure 23: Define the Sub-Subnets for Subnet #14 (140.25.224.0/20)

Since 16 = 2^4, four more bits are required to identify each of the 16 subnets. This means that the organization will need to use a /24 as the extended-network-prefix length.

The 16 subnets of the 140.25.224.0/20 address block are given below. The subnets are numbered 0 through 15. The underlined portion of each sub-subnet address identifies the extended-network-prefix, while the **bold** digits identify the 4-bits representing the sub-subnet-number field:

Subnet #14: 10001100.00011001.1110**0000**.00000000 = 140.25.224.0/20

Subnet #14-0: 10001100.00011001.1110**0000**.00000000 = 140.25.224.0/24
Subnet #14-1: 10001100.00011001.1110**0001**.00000000 = 140.25.225.0/24
Subnet #14-2: 10001100.00011001.1110**0010**.00000000 = 140.25.226.0/24
Subnet #14-3: 10001100.00011001.1110**0011**.00000000 = 140.25.227.0/24
Subnet #14-4: 10001100.00011001.1110**0100**.00000000 = 140.25.228.0/24
:
:
Subnet #14-14: 10001100.00011001.1110**1110**.00000000 = 140.25.238.0/24
Subnet #14-15: 10001100.00011001.1110**1111**.00000000 = 140.25.239.0/24

Define Host Addresses for Subnet #14-3 (140.25.227.0/24)

Let's examine the host addresses that can be assigned to Subnet #14-3 (140.25.227.0/24). This is illustrated in Figure 24.

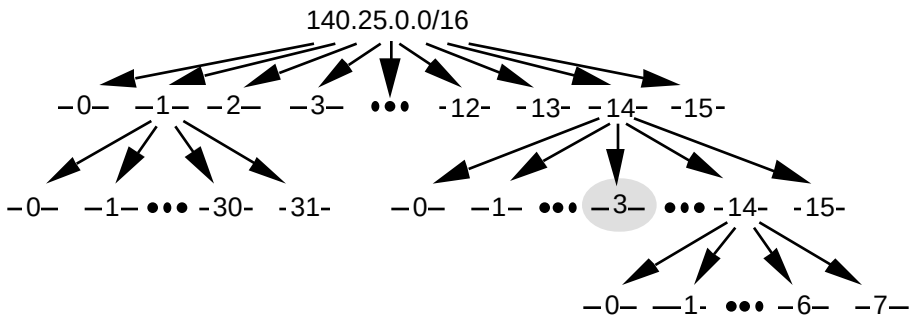

Figure 24: Define the Host Addresses for Subnet #14-3 (140.25.227.0/24)

Each of the subnets of Subnet #14-3 has 8 bits in the host-number field. This means that each subnet represents a block of 254 valid host addresses (2^8-2). The hosts are numbered 1 through 254.

The valid host addresses for Subnet #14-3 are given below. The underlined portion of each address identifies the extended-network-prefix, while the **bold** digits identify the 8-bit host-number field:

```
Subnet #14-3:  10001100.00011001.11100011.00000000 = 140.25.227.0/24

Host #1      10001100.00011001.11100011.00000001 = 140.25.227.1/24
Host #2      10001100.00011001.11100011.00000010 = 140.25.227.2/24
Host #3      10001100.00011001.11100011.00000011 = 140.25.227.3/24
Host #4      10001100.00011001.11100011.00000100 = 140.25.227.4/24
Host #5      10001100.00011001.11100011.00000101 = 140.25.227.5/24
    .
    .
Host #253    10001100.00011001.11100011.11111101 = 140.25.227.253/24
Host #254    10001100.00011001.11100011.11111110 = 140.25.227.254/24
```

The broadcast address for Subnet #14-3 is the all 1's host address or:

```
10001100.00011001.11100011.11111111 = 140.25.227.255
```

The broadcast address for Subnet #14-3 is exactly one less than the base address for Subnet #14-4 (140.25.228.0).

Define the Sub^2-Subnets for Subnet #14-14 (140.25.238.0/24)

After Subnet #14 was divided into sixteen subnets, Subnet #14-14 is further subdivided into 8 equal-size address blocks. This is illustrated in Figure 25.

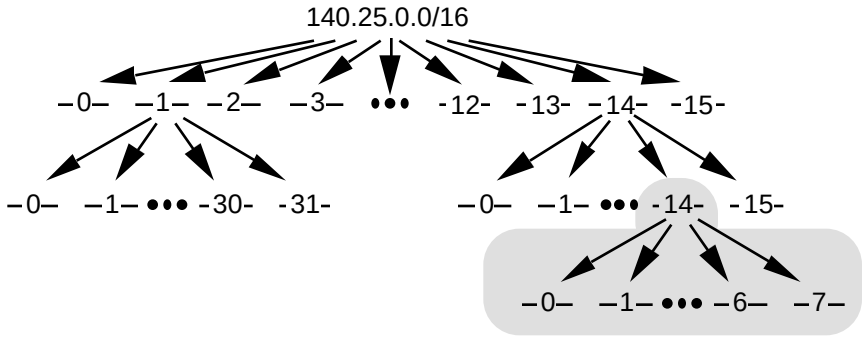

Figure 25: Define the Sub^2-Subnets for Subnet #14-14 (140.25.238.0/24)

Since $8 = 2^3$, three more bits are required to identify each of the 8 subnets. This means that the organization will need to use a /27 as the extended-network-prefix length.

The 8 subnets of the 140.25.238.0/24 address block are given below. The subnets are numbered 0 through 7. The underlined portion of each sub-subnet address identifies the extended-network-prefix, while the **bold** digits identify the 3-bits representing the subnet^2-number field:

```
Subnet #14-14: 10001100.00011001.11101110.00000000 = 140.25.238.0/24
```

```
Subnet#14-14-0:10001100.00011001.11101110.000000000 = 140.25.238.0/27
Subnet#14-14-1:10001100.00011001.11101110.001000000 = 140.25.238.32/27
Subnet#14-14-2:10001100.00011001.11101110.010000000 = 140.25.238.64/27
Subnet#14-14-3:10001100.00011001.11101110.011000000 = 140.25.238.96/27
Subnet#14-14-4:10001100.00011001.11101110.100000000 = 140.25.238.128/27
Subnet#14-14-5:10001100.00011001.11101110.101000000 = 140.25.238.160/27
Subnet#14-14-6:10001100.00011001.11101110.110000000 = 140.25.238.192/27
Subnet#14-14-7:10001100.00011001.11101110.111000000 = 140.25.238.224/27
```

Define Host Addresses for Subnet #14-14-2 (140.25.238.64/27)

Let's examine the host addresses that can be assigned to Subnet #14-14-2 (140.25.238.64/27). This is illustrated in Figure 26.

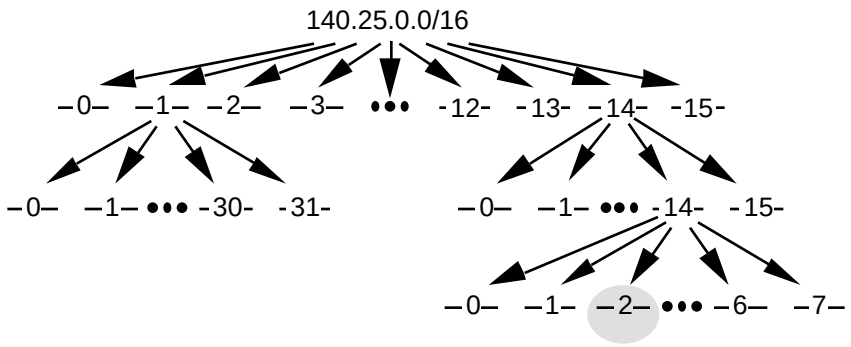

Figure 26: Define the Host Addresses for Subnet #14-14-2 (140.25.238.64/27)

Each of the subnets of Subnet #14-14 has 5 bits in the host-number field. This means that each subnet represents a block of 30 valid host addresses (2^5-2). The hosts will be numbered 1 through 30.

The valid host addresses for Subnet #14-14-2 are given below. The underlined portion of each address identifies the extended-network-prefix, while the **bold** digits identify the 5-bit host-number field:

```
Subnet#14-14-2:10001100.00011001.11101110.01000000 = 140.25.238.64/27

  Host #1     10001100.00011001.11101110.010000001 = 140.25.238.65/27
  Host #2     10001100.00011001.11101110.010000010 = 140.25.238.66/27
  Host #3     10001100.00011001.11101110.010000011 = 140.25.238.67/27
  Host #4     10001100.00011001.11101110.010000100 = 140.25.238.68/27
  Host #5     10001100.00011001.11101110.010000101 = 140.25.238.69/27
      .
      .
  Host #29    10001100.00011001.11101110.010011101 = 140.25.238.93/27
  Host #30    10001100.00011001.11101110.010011110 = 140.25.238.94/27
```

The broadcast address for Subnet #14-14-2 is the all 1's host address or:

```
10001100.00011001.11011100.010111111 = 140.25.238.95
```

The broadcast address for Subnet #6-14-2 is exactly one less than the base address for Subnet #14-14-3 (140.25.238.96).

Additional Practice with VLSM

Please turn to Appendix D for practice exerciss to reinforce your understanding of VLSM.

Classless Inter-Domain Routing (CIDR)

By 1992, the exponential growth of the Internet was beginning to raise serious concerns among members of the IETF about the ability of the Internet's routing system to scale and support future growth. These problems were related to:

- The near-term exhaustion of the Class B network address space
- The rapid growth in the size of the global Internet's routing tables
- The eventual exhaustion of the 32-bit IPv4 address space

Projected Internet growth figures made it clear that the first two problems were likely to become critical by 1994 or 1995. The response to these immediate challenges was the development of the concept of Supernetting or Classless Inter-Domain Routing (CIDR). The third problem, which is of a more long-term nature, is currently being explored by the IP Next Generation (IPng or IPv6) working group of the IETF.

CIDR was officially documented in September 1993 in RFC 1517, 1518, 1519, and 1520. CIDR supports two important features that benefit the global Internet routing system:

- CIDR eliminates the traditional concept of Class A, Class B, and Class C network addresses. This enables the efficient allocation of the IPv4 address space which will allow the continued growth of the Internet until IPv6 is deployed.
- CIDR supports route aggregation where a single routing table entry can represent the address space of perhaps thousands of traditional classful routes. This allows a single routing table entry to specify how to route traffic to many individual network addresses. Route aggregation helps control the amount of routing information in the Internet's backbone routers, reduces route flapping (rapid changes in route availability), and eases the local administrative burden of updating external routing information.

Without the rapid deployment of CIDR in 1994 and 1995, the Internet routing tables would have in excess of 70,000 routes (instead of the current 30,000+) and the Internet would probably not be functioning today!

CIDR Promotes the Efficient Allocation of the IPv4 Address Space

CIDR eliminates the traditional concept of Class A, Class B, and Class C network addresses and replaces them with the generalized concept of a "network-prefix." Routers use the network-prefix, rather than the first 3 bits of the IP address, to determine the dividing point between the network number and the host number. As a result, CIDR supports the deployment of *arbitrarily sized* networks rather than the standard 8-bit, 16-bit, or 24-bit network numbers associated with classful addressing.

In the CIDR model, each piece of routing information is advertised with a bit mask (or prefix-length). The prefix-length is a way of specifying the number of leftmost contiguous bits in the network-portion of each routing table entry. For example, a

network with 20 bits of network-number and 12-bits of host-number would be advertised with a 20-bit prefix length (a /20). The clever thing is that the IP address advertised with the /20 prefix could be a former Class A, Class B, or Class C. Routers that support CIDR do *not* make assumptions based on the first 3-bits of the address, they rely on the prefix-length information provided with the route.

In a classless environment, prefixes are viewed as bitwise contiguous blocks of the IP address space. For example, all prefixes with a /20 prefix represent the same amount of address space (2^{12} or 4,096 host addresses). Furthermore, a /20 prefix can be assigned to a traditional Class A, Class B, or Class C network number. Figure 27 shows how each of the following /20 blocks represent 4,096 host addresses - 10.23.64.0/20, 130.5.0.0/20, and 200.7.128.0/20.

```
Traditional A    10.23.64.0/20    00001010.00010111.01000000.00000000

Traditional B    130.5.0.0/20     10000010.00000101.00000000.00000000

Traditional C    200.7.128.0/20   11001000.00000111.10000000.00000000
```

Figure 27: /20 Bitwise Contiguous Address Blocks

Table 3 provides information about the most commonly deployed CIDR address blocks. Referring to the Table, you can see that a /15 allocation can also be specified using the traditional dotted-decimal mask notation of 255.254.0.0. Also, a /15 allocation contains a bitwise contiguous block of 128K (131,072) IP addresses which can be classfully interpreted as 2 Class B networks or 512 Class C networks.

Table 3: CIDR Address Blocks

CIDR prefix-length	Dotted-Decimal	# Individual Addresses	# of Classful Networks
/13	255.248.0.0	512 K	8 Bs or 2048 Cs
/14	255.252.0.0	256 K	4 Bs or 1024 Cs
/15	255.254.0.0	128 K	2 Bs or 512 Cs
/16	255.255.0.0	64 K	1 B or 256 Cs
/17	255.255.128.0	32 K	128 Cs
/18	255.255.192.0	16 K	64 Cs
/19	255.255.224.0	8 K	32 Cs
/20	255.255.240.0	4 K	16 Cs
/21	255.255.248.0	2 K	8 Cs
/22	255.255.252.0	1 K	4 Cs
/23	255.255.254.0	512	2 Cs
/24	255.255.255.0	256	1 C
/25	255.255.255.128	128	1/2 C
/26	255.255.255.192	64	1/4 C
/27	255.255.255.224	32	1/8 C

Host Implications for CIDR Deployment

It is important to note that there may be severe host implications when you deploy CIDR based networks. Since many hosts are classful, their user interface will not permit them to be configured with a mask that is shorter than the "natural" mask for a traditional classful address. For example, potential problems could exist if you wanted to deploy 200.25.16.0 as a /20 to define a network capable of supporting 4,094 (2^{12}-2) hosts. The software executing on each end station might not allow a traditional Class C (200.25.16.0) to be configured with a 20-bit mask since the natural mask for a Class C network is a 24-bit mask. If the host software supports CIDR, it will permit shorter masks to be configured.

However, there will be no host problems if you were to deploy the 200.25.16.0/20 (a traditional Class C) allocation as a block of 16 /24s since non-CIDR hosts will interpret their local /24 as a Class C. Likewise, 130.14.0.0/16 (a traditional Class B) could be deployed as a block of 255 /24s since the hosts will interpret the /24s as subnets of a /16. If host software supports the configuration of shorter than expected masks, the network manager has tremendous flexibility in network design and address allocation.

Efficient Address Allocation

How does all of this lead to the efficient allocation of the IPv4 address space? In a classful environment, an Internet Service Provider (ISP) can only allocate /8, /16, or /24 addresses. In a CIDR environment, the ISP can carve out a block of its registered address space that specifically meets the needs of each client, provides additional room for growth, and does not waste a scarce resource.

Assume that an ISP has been assigned the address block 206.0.64.0/18. This block represents 16,384 (2^{14}) IP addresses which can be interpreted as 64 /24s. If a client requires 800 host addresses, rather than assigning a Class B (and wasting ~64,700 addresses) or four individual Class Cs (and introducing 4 new routes into the global Internet routing tables), the ISP could assign the client the address block 206.0.68.0/22, a block of 1,024 (2^{10}) IP addresses (4 contiguous /24s). The efficiency of this allocation is illustrated in Figure 28.

ISP's Block:	11001110.00000000.01000000.00000000	206.0.64.0/18
Client Block:	11001110.00000000.01000100.00000000	206.0.68.0/22
Class C #0:	11001110.00000000.01000100.00000000	206.0.68.0/24
Class C #1:	11001110.00000000.01000101.00000000	206.0.69.0/24
Class C #2:	11001110.00000000.01000110.00000000	206.0.70.0/24
Class C #3:	11001110.00000000.01000111.00000000	206.0.71.0/24

Figure 28: CIDR Supports Efficient Address Allocation

CIDR Address Allocation Example

For this example, assume that an ISP owns the address block 200.25.0.0/16. This block represents 65, 536 (2^{16}) IP addresses (or 256 /24s).

From the 200.25.0.0/16 block it wants to allocate the 200.25.16.0/20 address block . This smaller block represents 4,096 (2^{12}) IP addresses (or 16 /24s).

```
Address Block   11001000.00011001.00010000.00000000   200.25.16.0/20
```

In a classful environment, the ISP is forced to use the /20 as 16 individual /24s.

```
Network #0      11001000.00011001.00010000.00000000   200.25.16.0/24
Network #1      11001000.00011001.00010001.00000000   200.25.17.0/24
Network #2      11001000.00011001.00010010.00000000   200.25.18.0/24
Network #3      11001000.00011001.00010011.00000000   200.25.19.0/24
Network #4      11001000.00011001.00010100.00000000   200.25.20.0/24
    :
    :
Network #13     11001000.00011001.00011101.00000000   200.25.29.0/24
Network #14     11001000.00011001.00011110.00000000   200.25.30.0/24
Network #15     11001000.00011001.00011111.00000000   200.25.31.0/24
```

If you look at the ISP's /20 address block as a pie, in a classful environment it can only be cut into 16 equal-size pieces. This is illustrated in Figure 29.

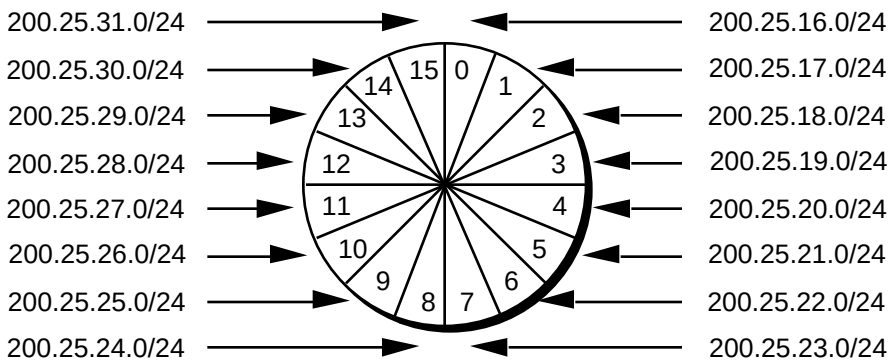

Figure 29: Slicing the Pie - Classful Environment

However, in a classless environment, the ISP is free to cut up the pie any way it wants. It could slice up the original pie into 2 pieces (each 1/2 of the address space) and assign one portion to Organization A, then cut the other half into 2 pieces (each 1/4 of the address space) and assign one piece to Organization B, and finally slice the remaining fourth into 2 pieces (each 1/8 of the address space) and assign it to Organization C and Organization D. Each of the individual organizations is free to allocate the address space within its "Intranetwork" as it sees fit. This is illustrated in Figure 30.

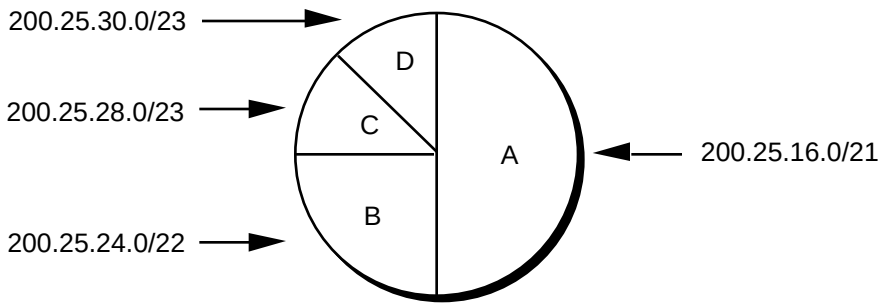

Figure 30: Slicing the Pie - Classless Environment

Step #1: Divide the address block 200.25.16.0/20 into two equal size slices. Each block represents one-half of the address space or 2,048 (2^{11}) IP addresses.

```
ISP's Block   11001000.00011001.00010000.00000000   200.25.16.0/20

Org A:        11001000.00011001.00010000.00000000   200.25.16.0/21
Reserved:     11001000.00011001.00011000.00000000   200.25.24.0/21
```

Step #2: Divide the reserved block (200.25.24.0/21) into two equal size slices. Each block represents one-fourth of the address space or 1,024 (2^{10}) IP addresses.

```
Reserved      11001000.00011001.00011000.00000000   200.25.24.0/21

Org B:        11001000.00011001.00011000.00000000   200.25.24.0/22
Reserved      11001000.00011001.00011100.00000000   200.25.28.0/22
```

Step #3: Divide the reserved address block (200.25.28.0/22) into two equal size blocks. Each block represents one-eight of the address space or 512 (2^{9}) IP addresses.

```
Reserved      11001000.00011001.00011100.00000000   200.25.28.0/22

Org C:        11001000.00011001.00011100.00000000   200.25.28.0/23
Org D:        11001000.00011001.00011110.00000000   200.25.30.0/23
```

CIDR is Similar to VLSM

If CIDR appears to have the familiar look and feel of VLSM, you're correct! CIDR and VLSM are essentially the same thing since they both allow a portion of the IP address space to be recursively divided into subsequently smaller pieces. The difference is that with VLSM, the recursion is performed on the address space previously assigned to an organization and is invisible to the global Internet. CIDR, on the other hand, permits the recursive allocation of an address block by an Internet Registry to a high-level ISP, to a mid-level ISP, to a low-level ISP, and finally to a private organization's network.

Just like VLSM, the successful deployment of CIDR has three prerequisites:

- The routing protocols must carry network-prefix information with each route advertisement.
- All routers must implement a consistent forwarding algorithm based on the "longest match."
- For route aggregation to occur, addresses must be assigned so that they are topologically significant.

Controlling the Growth of Internet's Routing Tables

Another important benefit of CIDR is that it plays an important role in controlling the growth of the Internet's routing tables. The reduction of routing information requires that the Internet be divided into addressing domains. Within a domain, detailed information is available about all of the networks that reside in the domain. Outside of an addressing domain, only the common network prefix is advertised. This allows a single routing table entry to specify a route to many individual network addresses.

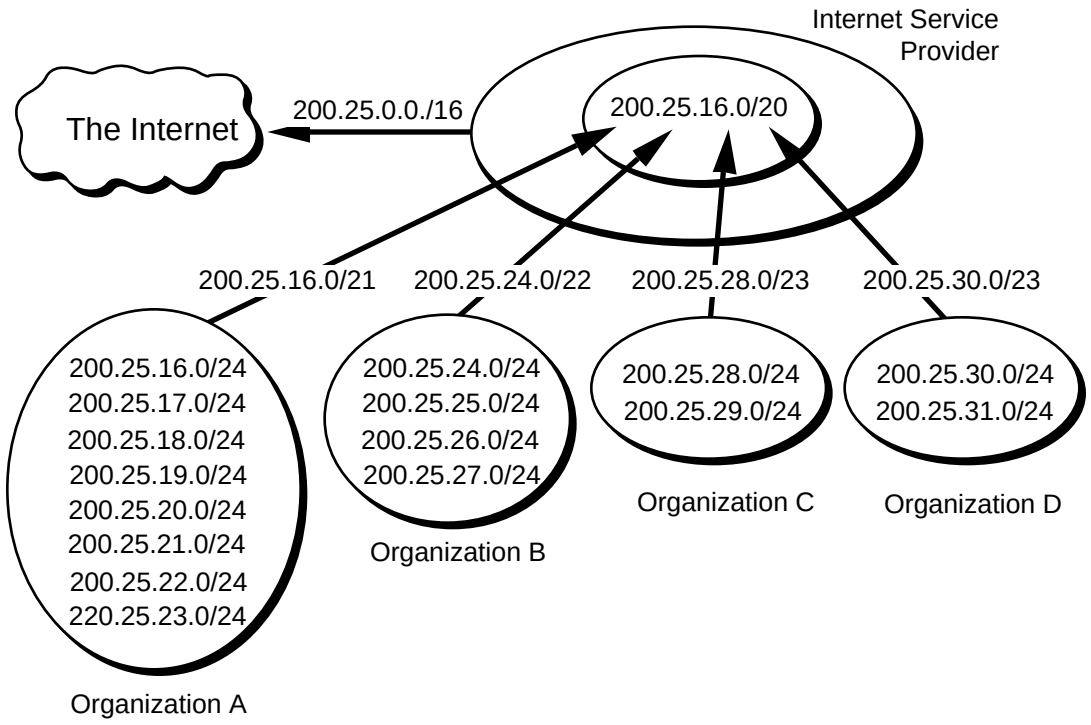

Figure 31: CIDR Reduces the Size of Internet Routing Tables

Figure 31 illustrates how the allocation described in previous CIDR example helps reduce the size of the Internet routing tables. Assume that a portion of the ISPs address block (200.25.16.0/20) has been allocated as described in the previous example. Organization A aggregates 8 /24s into a single advertisement (200.25.16.0/21), Organization B aggregates 4 /24s into a single advertisement (200.25.24.0/22), Organization C aggregates 2 /24s into a single advertisement (200.25.28.0/23), and

Organization D aggregates 2 /24s into a single advertisement (200.25.30.0/23). Finally, the ISP is able to inject the 256 /24s in its allocation into the Internet with a single advertisement - 200.25.0.0/16!

It should be mentioned that route aggregation via BGP-4 is not automatic. The network engineers must configure each router to perform the required aggregation. The successful deployment of CIDR will allow the number of individual networks on the Internet to expand, while minimizing the number of routes in the Internet routing tables.

Routing in a Classless Environment

Figure 32 illustrates the routing advertisements for Organization A discussed in the previous CIDR Example.

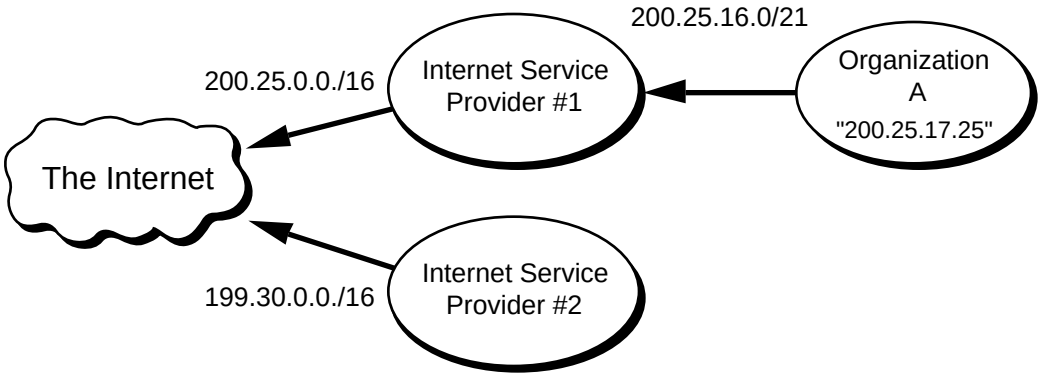

Figure 32: Routing Advertisements for Organization A

Since all of Organization A's routes are part of ISP #1's address block, the routes to Organization A are implicitly aggregated via ISP #1's aggregated announcement to the Internet. In other words, the eight networks assigned to Organization A are hidden behind a single routing advertisement. Using the longest match forwarding algorithm, Internet routers will route traffic to host 200.25.17.25 to ISP #1, which will in turn route the traffic to Organization A.

Now, for whatever reasons, assume that Organization A decides to change its network provider to ISP #2. This is illustrated in Figure 33.

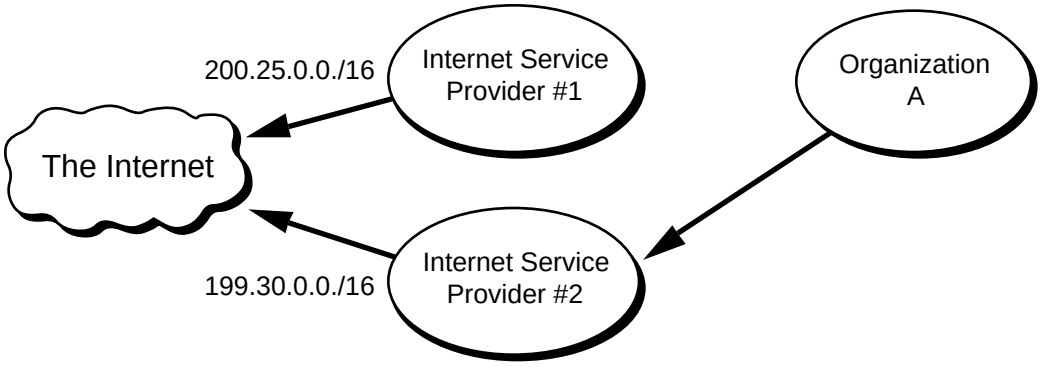

Figure 33: Organization A Changes Network Providers to ISP #2

The "best" thing for the size of the Internet's routing tables would be to have Organization A obtain a block of ISP #2's address space and renumber. This would allow the eight networks assigned to Organization A to be hidden behind the aggregate routing advertisement of ISP #2. Unfortunately, renumbering is a labor-intensive task which could be very difficult, if not impossible, for Organization A.

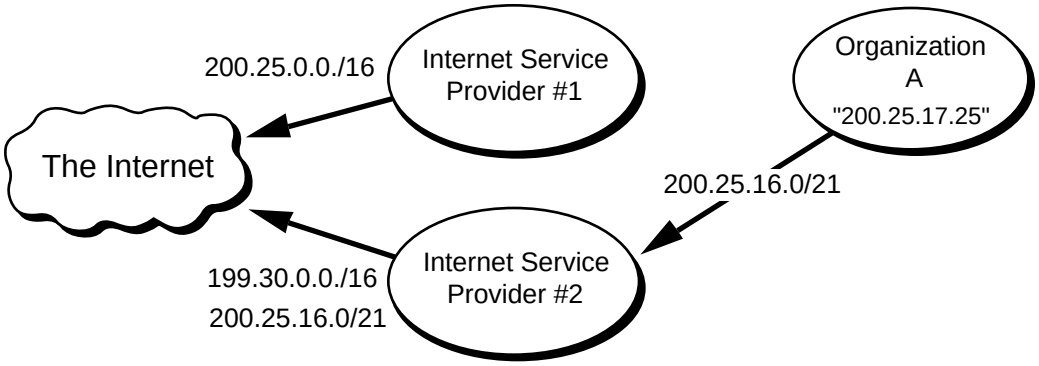

Figure 34: ISP #2 Injects a More-Specific Route into the Internet

The "best" thing for Organization A is to retain ownership of its address space and have ISP #2 advertise an "exception" (more specific) route into the Internet. The exception route allows all traffic for 200.25.0.0/16 to be sent to ISP #1, with the exception of the traffic to 200.25.16.0/21. This is accomplished by having ISP #2 advertise, in addition to its own 199.30.0.0/16 block, a route for 200.25.16.0/21. Please refer to Figure 34. Using the "longest match" forwarding algorithm, Internet routers will route traffic addressed to host 200.25.17.25 to ISP #2 which will in turn route the traffic to Organization A. Clearly, the introduction of a large number of exception routes can reduce the effectiveness of the CIDR deployment and eventually cause Internet routing tables to begin exploding again!

NETBuilder Support for CIDR

Support for CIDR has been implemented on the NETBuilder:

- NETBuilder software implements BGP-4. Support for CIDR is a significant part of the improvements made to BGP-4.
- NETBuilder software uses a routing table structure that understands a network number advertised with a prefix that is shorter than the natural mask. The NETBuilder's routing table and forwarding process ignore the traditional IP address Class and are capable of accepting any network/mask combination that it receives.
- NETBuilder software is capable of performing aggregation by way of BGP-4 configuration parameters. Also, the OSPF AreaRange parameter allows VLSM-based aggregation to be performed within an autonomous system. The network administrator may specify exactly what network numbers and masks are advertised outside of each area or domain.

Additional Practice with CIDR

Please turn to Appendix E for several practice exercises to reinforce your understanding of CIDR.

New Solutions for Scaling the Internet Address Space

As we approach the turn of the century, the problems of IPv4 address shortages and expanding Internet routing tables are still with us. The good news is that CIDR is working. The bad news is that recent growth trends indicate that the number of Internet routes is beginning to, once again, increase at an exponential rate. The Internet must find a way to keep the routing table growth linear. The IETF is continuing its efforts to develop solutions that will overcome these problems, enabling the continued growth and scalability of the Internet.

Appeal to Return Unused IP Network Prefixes

RFC 1917 requests that the Internet community return unused address blocks to the Internet Assigned Numbers Authority (IANA) for redistribution. This includes unused network numbers, addresses for networks that will never be connected to the global Internet for security reasons, and sites that are using a small percentage of their address space. RFC 1917 also petitions ISPs to return unused network-prefixes that are outside of their assigned address blocks. It will be interesting to see how the Internet community responds since many organizations with unused addresses don't want to return them because they are viewed as an asset.

Address Allocation for Private Internets

RFC 1918 requests that organizations make use of the private Internet address space for hosts that require IP connectivity within their enterprise network, but do not require external connections to the global Internet. For this purpose, the IANA has reserved the following three address blocks for private internets:

```
10.0.0.0    - 10.255.255.255      (10/8 prefix)
172.16.0.0  - 172.31.255.255      (172.16/12 prefix)
192.168.0.0 - 192.168.255.255     (192.168/16 prefix)
```

Any organization that elects to use addresses from these reserved blocks can do so without contacting the IANA or an Internet registry. Since these addresses are never injected into the global Internet routing system, the address space can simultaneously be used by many different organizations.

The disadvantage to this addressing scheme is that it requires an organization to use a Network Address Translator (NAT) for global Internet access. However, the use of the private address space and a NAT make it much easier for clients to change their ISP without the need to renumber or "punch holes" in a previously aggregated advertisement. The benefits of this addressing scheme to the Internet is that it reduces the demand for IP addresses so large organizations may require only a small block of the globally unique IPv4 address space.

Address Allocation from the Reserved Class A Address Space

An Internet draft, "Observations on the use of Components of the Class A Address Space within the Internet" <draft-ietf-cidrd-classa-01.txt>, explores the allocation of the upper-half of the currently reserved Class A address space through delegated registries. As the demand for IP addresses continues to grow, it appears that it may be necessary to eventually allocate the 64.0.0.0/2 address space. Note that the 64.0.0.0/2 address block is huge and represents 25% of the IPv4 unicast address space.

Implications of Address Allocation Policies

An Internet draft , "Implications of Various Address Allocation Policies for Internet Routing" <draft-ietf-cidrd-addr-ownership-07.txt>, discusses the fundamental issues that must be considered as the Internet develops a new unicast address allocation and management policies. The draft compares the benefits and limitations of an "address ownership" policy with an "address lending" policy.

"Address ownership" means that when an address block is assigned to an organization, it remains allocated to that organization for as long as the organization wants to keep it. This means that the address block is "portable" and that the organization would be able to use it to gain access to the Internet no matter where the organization connects to the Internet. On the other hand, "address lending" means that an organization obtains its address block on a "loan" basis. If the loan ends, the organization can no longer use the borrowed address block, must obtain new addresses, and renumber before using them.

As we have seen, hierarchical routing requires that addresses reflect the network topology in order to permit route aggregation. The draft argues that there are two fundamental problems that break the hierarchical addressing and routing model supported by CIDR:

- The continued existence of pre-CIDR routes that cannot be aggregated.
- Organizations that switch ISPs and continue to use addresses from their previous ISP's address block. The new ISP cannot aggregate the old address block as part of its aggregation, so it must inject an exception route into the Internet. If the number of exception routes continues to increases, they will erode the benefits of CIDR and prevent the scalability of the Internet's routing system.

The draft concludes with the recommendation that large providers, which can express their destinations with a single prefix, be assigned address blocks following the "address ownership" model. However, all allocations from these providers to a downstream clients should follow the "address lending" model. This means that if an organization changes its provider, the loan is canceled and the client will be required to renumber.

This draft has generated a tremendous amount of discussion within the Internet community about the concept of address ownership and what it means in the context of global routing. The authors present a strong argument that the Internet has to make a choice between either address ownership for all or a routable Internet - it can't have

both! Smaller organizations that want to own their addresses have concerns about the difficulty of renumbering and their lack of self-determination if their provider or their provider's upstream provider changes its provider. Finally, ISPs have concerns because the term "large provider" has not been defined. At this time, the discussion continues since any criteria recommended by the IETF is bound to be perceived as unfair by some!

Procedures for Internet/Enterprise Renumbering (PIER)

In the face of the "address ownership" vs. "address lending" debate, it is clear that renumbering may become a critical issue in the late 1990s. Procedures for Internet/Enterprise Renumbering (PIER) is a working group of the IETF charged with the task of developing a renumbering strategy.

RFC 1916 is a request by PIER for the Internet community to provide assistance in the development of a series of documents describing how an organization might proceed to renumber its network. The ultimate goal of these documents is to provide education and practical experience to the Internet community.

Market-Based Allocation of IP Address Blocks

An Internet draft ,"Suggestions for Market-Based Allocation of IP Address Blocks" <draft-ietf-cidrd-blocks-00.txt>, is a proposal to make IPv4 address assignments transferable and condones the exchange of money as part of the transfer procedure. It suggests that the Internet community embrace the profit motive as an incentive to motivate organizations to act in ways that will improve resource use. This proposal goes hand-in-hand with another proposal to introduce financial incentives for route aggregation (i.e., have ISPs levy a charge for each route advertised). The idea is to move the decisions regarding scarce resources from a political atmosphere to a financial environment which is better suited to deal with scarcity.

Keeping Current on Internet Addressing Issues

General Internet Information

Internet Monthly Reports discuss the accomplishments, milestones, and problems discovered on the Internet. They are available from: <http://info.internet.isi.edu/1/in-notes/imr>

Minutes of the most recent IETF Proceedings are available from: <http://www.ietf.cnri.reston.va.us/proceedings/directory.html>

Information about the size and content of the Internet routing table is available on the Merit Web pages: <http://www.ra.net/~ra/statistics/routes.html>

CIDR Deployment (CIDRD)

For general information about the CIDRD working group of the IETF and its charter: <http://www.ietf.cnri.reston.va.us/html.charters/cidrd-charter.html>

To subscribe to the CIDRD mailing list: <cidrd-request@iepg.org>