CHAPTER 2

Designing Large-Scale IP Internetworks

This chapter focuses on the following design implications of the Enhanced Interior Gateway Routing Protocol (IGRP) and Open Shortest Path First (OSPF) protocols:

- Network Topology
- Addressing and Route Summarization
- Route Selection
- Convergence
- Network Scalability
- Security

Enhanced IGRP and OSPF are routing protocols for the Internet Protocol (IP). An introductory discussion outlines general routing protocol issues; subsequent discussions focus on design guidelines for the specific IP protocols.

Implementing Routing Protocols

The following discussion provides an overview of the key decisions you must make when selecting and deploying routing protocols. This discussion lays the foundation for subsequent discussions regarding specific routing protocols.

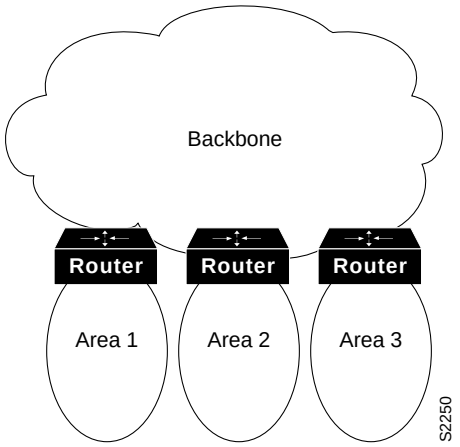
Network Topology

The physical topology of an internetwork is described by the complete set of routers and the networks that connect them. Networks also have a logical topology. Different routing protocols establish the logical topology in different ways.

Some routing protocols do not use a logical hierarchy. Such protocols use addressing to segregate specific areas or domains within a given internetworking environment and to establish a logical topology. For such nonhierarchical, or *flat*, protocols, no manual topology creation is required.

Other protocols require the creation of an explicit hierarchical topology through establishment of a backbone and logical areas. The OSPF and Intermediate System-to-Intermediate System (IS-IS) protocols are examples of routing protocols that use a hierarchical structure. A general hierarchical network scheme is illustrated in Figure 2-1. The explicit topology in a hierarchical scheme takes precedence over the topology created through addressing.

Figure 2-1 Hierarchical Network



If a hierarchical routing protocol is used, the addressing topology should be assigned to reflect the hierarchy. If a flat routing protocol is used, the addressing implicitly creates the topology.

There are two recommended ways to assign addresses in a hierarchical network. The simplest way is to give each area (including the backbone) a unique network address. An alternative is to assign address ranges to each area.

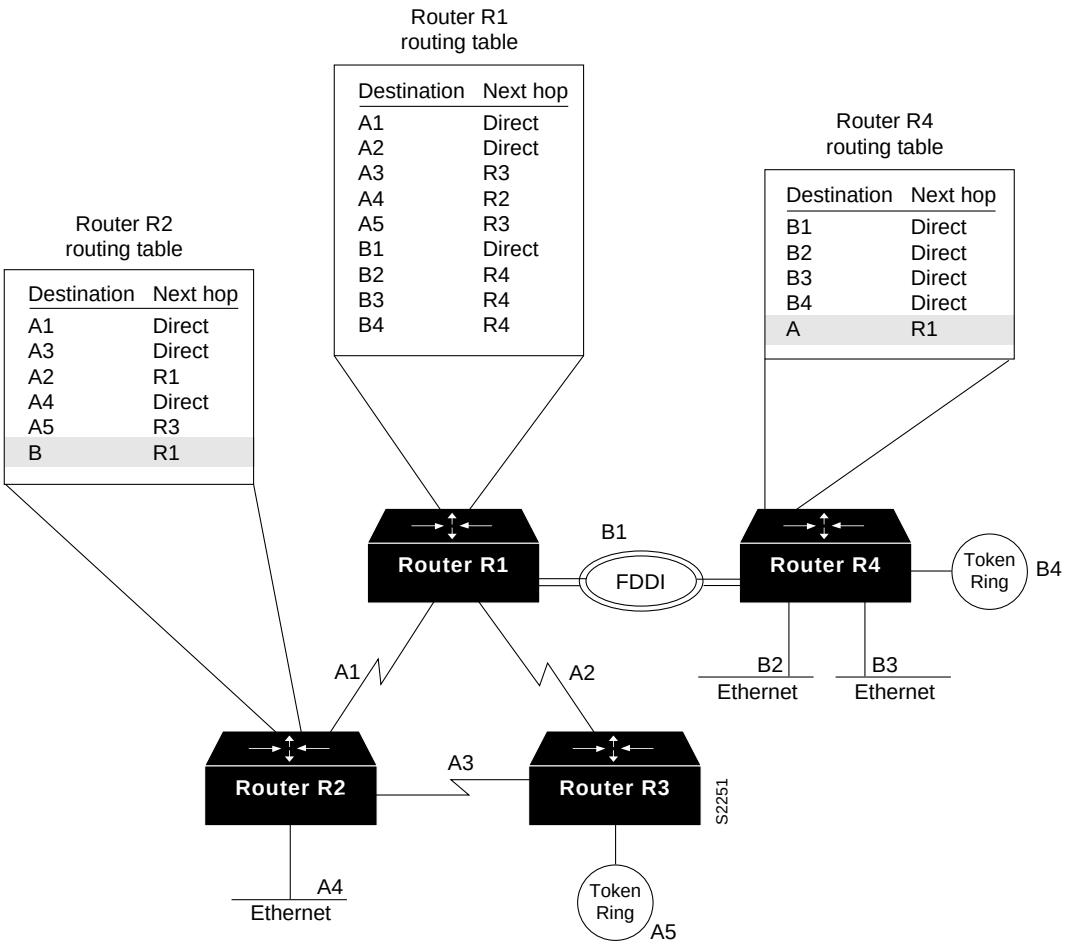
Areas are logical collections of contiguous networks and hosts. Areas also include all the routers having interfaces on any one of the included networks. Each area runs a separate copy of the basic routing algorithm. Therefore, each area has its own topological database.

Addressing and Route Summarization

Route summarization procedures condense routing information. Without summarization, each router in a network must retain a route to every subnet in the network. With summarization, routers can reduce some sets of routes to a single advertisement, reducing both the load on the router and the perceived complexity of the network. The importance of route summarization increases with network size.

Figure 2-2 illustrates an example of route summarization. In this environment, Router R2 maintains one route for all destination networks beginning with “B” and Router R4 maintains one route for all destination networks beginning with “A.” This is the essence of route summarization. Router R1 tracks all routes because it exists on the boundary between “A” and “B.”

Figure 2-2 Route Summarization Example

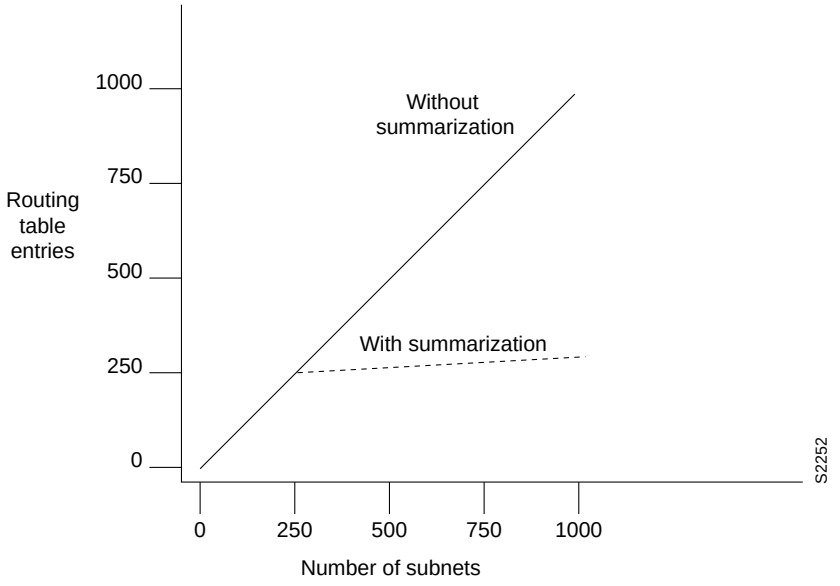


The reduction in route propagation and routing information overhead can be significant. Figure 2-3 illustrates the potential savings. The vertical axis of Figure 2-3 shows the number of routing table entries. The horizontal axis measures the number of subnets. Without summarization, each router in a network with 1000 subnets must contain 1000 routes. With summarization, the picture changes considerably. If you assume a Class B network with 8 bits of subnet address space, each router needs to know all of the routes for each subnet in its network number (250 routes, assuming that 1000 subnets fall into four major networks of 250 routers each) plus one route for each of the other networks (3) for a total of 253 routes. This represents a nearly 75-percent reduction in the size of the routing table.

The preceding example shows the simplest type of route summarization: collapsing all the subnet routes into a single network route. Some routing protocols also support route summarization at any bit boundary (rather than just at major network number boundaries) in a network address. A routing protocol can summarize on a bit boundary only if it supports *variable-length subnet masks* (VLSMs).

Some routing protocols summarize automatically. Other routing protocols require manual configuration to support route summarization.

Figure 2-3 Route Summarization Benefits

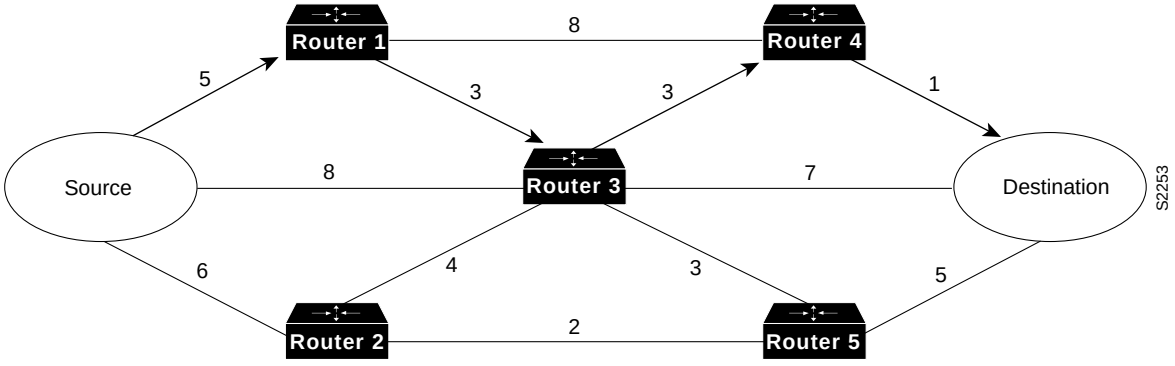


Route Selection

Route selection is trivial when only a single path to the destination exists. However, if any part of that path should fail, there is no way to recover. Therefore, most networks are designed with multiple paths so there are alternatives in case a failure occurs.

Routing protocols compare route metrics to select the best route from a group of possible routes. Route metrics are computed by assigning a characteristic or set of characteristics to each physical network. The metric for the route is an aggregation of the characteristics of each physical network in the route. Figure 2-4 shows a typical meshed network with metrics assigned to each link and the best route from source to destination identified.

Figure 2-4 Routing Metrics and Route Selection



Routing protocols use different techniques for assigning metrics to individual networks. Further, each routing protocol forms a metric aggregation in a different way.

Most routing protocols can use multiple paths if the paths have an equal cost. Some routing protocols can even use multiple paths when paths have an unequal cost. In either case, load balancing can improve overall allocation of network bandwidth.

When multiple paths are used, there are several ways to distribute the packets. The two most common mechanisms are per-packet load balancing and per-destination load balancing. Per-packet load balancing distributes the packets across the possible routes in a manner proportional to the route metrics. With equal-cost routes, this is equivalent to a round-robin scheme. One packet or destination (depending on switching mode) is distributed to each possible path. Per-destination load balancing distributes packets across the possible routes based on destination. Each new destination is assigned the next available route. This technique tends to preserve packet order.

Note Most TCP implementations can accommodate out-of-order packets. However, out-of-order packets may cause performance degradation.

When fast switching is enabled on a router (default condition), route selection is done on a per-destination basis. When fast switching is disabled, route selection is done on a per-packet basis. For line speeds of 56 kbps and faster, fast switching is recommended.

Convergence

When *network* topology changes, network traffic must reroute quickly. The phrase “convergence time” describes the time it takes a router to start using a new route after a topology changes.

Routers must do three things after a topology changes:

- Detect the change
- Select a new route
- Propagate the changed route information

Some changes are immediately detectable. For example, serial line failures that involve carrier loss are immediately detectable by a router. Other failures are harder to detect. For example, if a serial line becomes unreliable but carrier is not lost, the unreliable link is not immediately detectable. In addition, some media (Ethernet, for example) do not provide physical indications such as carrier loss. When a router is reset, other routers do not detect this immediately. In general, failure detection is dependent on the media involved and the routing protocol used.

Once a failure has been detected, the routing protocol must select a new route. The mechanisms used to do this are protocol dependent. All routing protocols must propagate the changed route. The mechanisms used to do this are also protocol dependent.

Network Scalability

The ability to extend your internetwork is determined, in part, by the scaling characteristics of the routing protocols used and the quality of the network design.

Network scalability is limited by two factors: operational issues and technical issues. Typically, operational issues are more significant than technical issues. Operational scaling concerns encourage the use of large areas or protocols that do not require hierarchical structures. When hierarchical protocols are required, technical scaling concerns promote the use of small areas. Finding the right balance is the art of network design.

From a technical standpoint, routing protocols scale well if their resource use grows less than linearly with the growth of the network. Three critical resources are used by routing protocols: memory, central processing unit (CPU), and bandwidth.

Memory

Routing protocols use memory to store routing tables and topology information. Route summarization cuts memory consumption for all routing protocols. Keeping areas small reduces the memory consumption for hierarchical routing protocols.

CPU

CPU usage is protocol dependent. Some protocols use CPU cycles to compare new routes to existing routes. Other protocols use CPU cycles to regenerate routing tables after a topology change. In most cases, the latter technique will use more CPU cycles than the former. For link-state protocols, keeping areas small and using summarization reduces CPU requirements by reducing the effect of a topology change and by decreasing the number of routes that must be recomputed after a topology change.

Bandwidth

Bandwidth usage is also protocol dependent. Three key issues determine the amount of bandwidth a routing protocol consumes:

- When routing information is sent—Periodic updates are sent at regular intervals. Flash updates are sent only when a change occurs.
- What routing information is sent—Complete updates contain all routing information. Partial updates contain only changed information.
- Where routing information is sent—Flooded updates are sent to all routers. Bounded updates are sent only to routers that are affected by a change.

Note These three issues also affect CPU usage.

Distance vector protocols such as Routing Information Protocol (RIP), Interior Gateway Routing Protocol (IGRP), Internetwork Packet Exchange (IPX) RIP, IPX Service Advertisement Protocol (SAP), and Routing Table Maintenance Protocol (RTMP), broadcast their complete routing table periodically, regardless of whether the routing table has changed. This periodic advertisement varies from every 10 seconds for RTMP to every 90 seconds for IGRP. When the network is stable, distance vector protocols behave well but waste of bandwidth because of the periodic sending of routing table updates, even when no change has occurred. When a failure occurs in the network, distance vector protocols do not add excessive load to the network, but they take a long time to reconverge to an alternate path or to flush a bad path from the network.

Link-state routing protocols, such as Open Shortest Path First (OSPF), Intermediate System-to-Intermediate System (IS-IS), and NetWare Link Services Protocol (NLSP), were designed to address the limitations of distance vector routing protocols (slow convergence and unnecessary bandwidth usage). Link-state protocols are more complex than distance vector protocols, and running them adds to the router's overhead. The additional overhead (in the form of memory utilization and bandwidth consumption when link-state protocols first start up) constrains the number of neighbors that a router can support and the number of neighbors that can be in an area.

When the network is stable, link-state protocols minimize bandwidth usage by sending updates only when a change occurs. A hello mechanism ascertains reachability of neighbors. When a failure occurs in the network, link-state protocols flood link-state advertisements (LSAs) throughout an area. LSAs cause every router within the failed area to recalculate routes. The fact that LSAs need to be flooded throughout the area in failure mode and the fact that all routers recalculate routing tables constrain the number of neighbors that can be in an area.

Enhanced IGRP is an advanced distance vector protocol that has some of the properties of link-state protocols. Enhanced IGRP addresses the limitations of conventional distance vector routing protocols (slow convergence and high bandwidth consumption in a steady state network). When the network is stable, Enhanced IGRP sends updates only when a change in the network occurs. Like link-state protocols, Enhanced IGRP uses a hello mechanism to determine the reachability of neighbors. When a failure occurs in the network, Enhanced IGRP looks for feasible successors by sending messages to its neighbors. The search for feasible successors can be aggressive in terms of the traffic it generates (updates, queries and replies) to achieve convergence. This behavior constrains the number of neighbors that are possible.

In WANs, consideration of bandwidth is especially critical. For example, Frame Relay, which statistically multiplexes many logical data connections (virtual circuits) over a single physical link, allows the creation of networks that share bandwidth. Public Frame Relay networks use bandwidth sharing at all levels within the network. That is, bandwidth sharing may occur within the Frame Relay network of Corporation X, as well as between the networks of Corporation X and Corporation Y.

Two factors have a substantial effect on the design of public Frame Relay networks:

- Users are charged for each permanent virtual circuit (PVC), which encourages network designers to minimize the number of PVCs.
- Public carrier networks sometimes provide incentives to avoid the use of committed information rate (CIR) circuits. Although service providers try to ensure sufficient bandwidth, packets can be dropped.

Overall, WANs can lose packets because of lack of bandwidth. For Frame Relay networks, this possibility is compounded because Frame Relay does not have a broadcast replication facility, so for every broadcast packet that is sent out a Frame Relay interface, the router must replicate it for each PVC on the interface. This requirement limits the number of PVCs that a router can handle effectively.

In addition to bandwidth, network designers must consider the size of routing tables that need to be propagated. Clearly, the design considerations for an interface with 50 neighbors and 100 routes to propagate are very different from the considerations for an interface with 50 neighbors and 10,000 routes to propagate. Table 2-1 gives a rough estimate of the number of WAN neighbors that a routing protocol can handle effectively.

Table 2-1 Routing Protocols and Number of WAN Neighbors

Routing Protocol	Number of Neighbors per Router
Distance vector	50
Link state	30
Advanced distance vector	30

Security

Controlling access to network resources is a primary concern. Some routing protocols provide techniques that can be used as part of a security strategy.

With some routing protocols, you can insert a filter on the routes being advertised so that certain routes are not advertised in some parts of the network.

Some routing protocols can authenticate routers that run the same protocol. Authentication mechanisms are protocol specific and generally weak. In spite of this, it is worthwhile to take advantage of the techniques that exist. Authentication can increase network stability by preventing unauthorized routers or hosts from participating in the routing protocol, whether those devices are attempting to participate accidentally or deliberately.

Enhanced IGRP Internetwork Design Guidelines

The Enhanced Interior Gateway Routing Protocol (Enhanced IGRP) is a routing protocol developed by Cisco Systems and introduced with Software Release 9.21 and Cisco Internetworking Operating System (Cisco IOS) Software Release 10.0. Enhanced IGRP combines the advantages of distance vector protocols, such as IGRP, with the advantages of link-state protocols, such as Open Shortest Path First (OSPF). Enhanced IGRP uses the Diffusing Update ALgorithm (DUAL) to achieve convergence quickly.

Enhanced IGRP includes support for IP, Novell NetWare, and AppleTalk. The discussion on Enhanced IGRP covers the following topics:

- Enhanced IGRP Network Topology
- Enhanced IGRP Addressing
- Enhanced IGRP Route Summarization
- Enhanced IGRP Route Selection
- Enhanced IGRP Convergence
- Enhanced IGRP Network Scalability
- Enhanced IGRP Security

Note Although the general discussion in this section is applicable to IP, IPX, and AppleTalk Enhanced IGRP, IP issues are highlighted here. For case studies on how to integrate Enhanced IGRP into IP, IPX, and AppleTalk networks, including detailed configuration examples and protocol-specific issues, see Chapter 1, “Integrating Enhanced IGRP into Existing Networks” in the Cisco publication *Internetworking Case Studies*.

Caution If you are using *candidate default route* in IP Enhanced IGRP and have installed multiple releases of Cisco router software within your internetwork that include any versions prior to September 1994, contact your Cisco technical support representative for version compatibility and software upgrade information. Refer to your software release notes for details.

Enhanced IGRP Network Topology

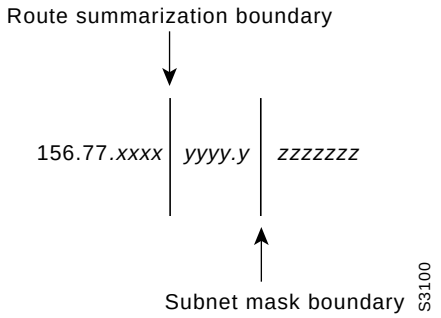
Enhanced IGRP uses a nonhierarchical (or flat) topology by default. Enhanced IGRP automatically summarizes subnet routes of directly connected networks at a network number boundary. This automatic summarization is sufficient for most IP networks. See the section “Enhanced IGRP Route Summarization” later in this chapter for more detail.

Enhanced IGRP Addressing

The first step in designing an Enhanced IGRP network is to decide on how to address the network. In many cases, a company is assigned a single NIC address (such as a Class B network address) to be allocated in a corporate internetwork. Bit-wise subnetting and variable-length subnetwork masks (VLSMs) can be used in combination to save address space. Enhanced IGRP for IP supports the use of VLSMs.

Consider a hypothetical network where a Class B address is divided into subnetworks, and contiguous groups of these subnetworks are summarized by Enhanced IGRP. The Class B network 156.77.0.0 might be subdivided as illustrated in Figure 2-5.

Figure 2-5 Variable-Length Subnet Masks (VLSMs) and Route Summarization Boundaries



In Figure 2-5, the letters x, y, and z represent bits of the last two octets of the Class B network as follows:

- The four *x* bits represent the route summarization boundary.
- The five *y* bits represent up to 32 subnets per summary route.
- The seven *z* bits allow for 126 (128-2) hosts per subnet.

Appendix A, “Subnetting an IP Address Space,” provides a complete example illustrating assignment for the Class B address 150.100.0.0.

Enhanced IGRP Route Summarization

With Enhanced IGRP, subnet routes of directly connected networks are automatically summarized at network number boundaries. In addition, a network administrator can configure route summarization at any interface with any bit boundary, allowing ranges of networks to be summarized arbitrarily.

Enhanced IGRP Route Selection

Routing protocols compare route metrics to select the best route from a group of possible routes. The following factors are important to understand when designing an Enhanced IGRP internetwork.

Enhanced IGRP uses the same vector of metrics as IGRP. Separate metric values are assigned for bandwidth, delay, reliability and load. By default, Enhanced IGRP computes the metric for a route by using the minimum bandwidth of each hop in the path and adding a media-specific delay for each hop. The metrics used by Enhanced IGRP are as follows:

- Bandwidth

  Bandwidth is deduced from the interface type. Bandwidth can be modified with the **bandwidth** command.

- Delay

  Each media type has a propagation delay associated with it. Modifying delay is very useful to optimize routing in network with satellite links. Delay can be modified with the **delay** command.

- Reliability

  Reliability is dynamically computed as a rolling weighted average over five seconds.

- Load

  Load is dynamically computed as a rolling weighted average over five seconds.

When Enhanced IGRP summarizes a group of routes, it uses the metric of the best route in the summary as the metric for the summary.

Note For information on Enhanced IGRP load sharing, see the section "IP Routing Protocols with Parallel Links" in Chapter 3, "Designing SRB Internetworks."

Enhanced IGRP Convergence

Enhanced IGRP implements a new convergence algorithm known as DUAL (Diffusing Update ALgorithm). DUAL uses two techniques that allow Enhanced IGRP to converge very quickly. First, each Enhanced IGRP router stores its neighbors routing tables. This allows the router to use a new route to a destination instantly if another *feasible* route is known. If no feasible route is known based upon the routing information previously learned from its neighbors, a router running Enhanced IGRP becomes *active* for that destination and sends a query to each of its neighbors asking for an alternate route to the destination. These queries propagate until an alternate route is found. Routers that are not affected by a topology change remain *passive* and do not need to be involved in the query and response.

A router using Enhanced IGRP receives full routing tables from its neighbors when it first communicates with the neighbors. Thereafter, only *changes* to the routing tables are sent and only to *routers* that are *affected* by the change. A *successor* is a neighboring router that is currently being used for packet forwarding, provides the *least cost* route to the destination, and is not part of a routing loop. Information in the routing table is based on *feasible successors.* Feasible successor routes can be used in case the existing route fails. Feasible successors provide the *next least-cost* path without introducing routing loops.

The routing table keeps a list of the computed costs of reaching networks. The topology table keeps a list of all routes advertised by neighbors. For each network, the router keeps the real cost of getting to that network and also keeps the advertised cost from its neighbor. In the event of a failure,

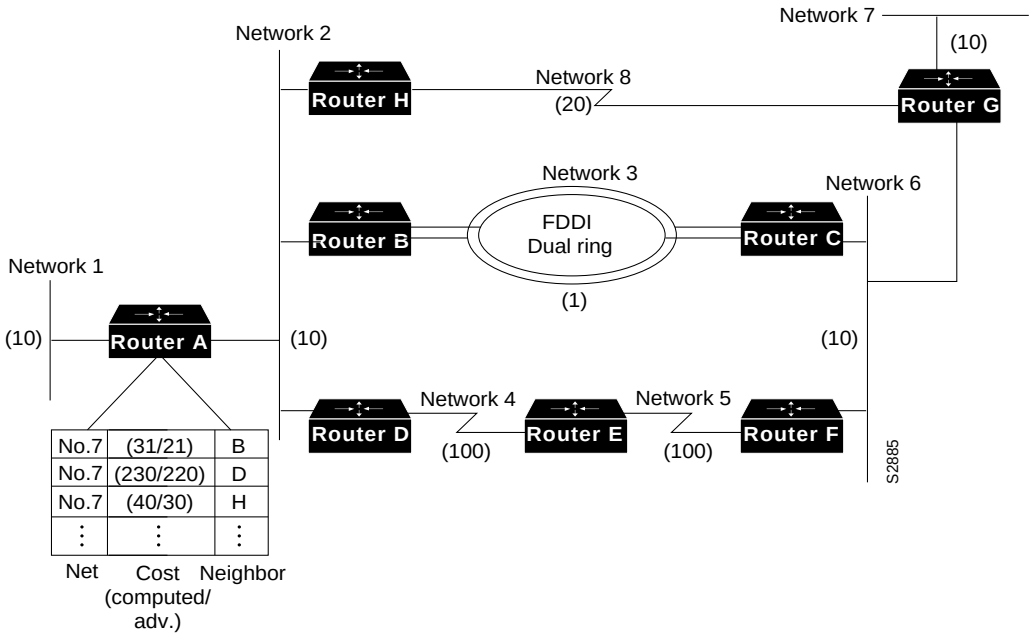
convergence is instant if a feasible successor can be found. A neighbor is a feasible successor if it meets the feasibility condition set by DUAL. DUAL finds feasible successors by the performing the following computations:

1. Determines membership of V_1. V_1 is the set of all neighbors whose advertised distance to network *x* is less than FD. (FD is the feasible distance and is defined as the best metric during an active-to-passive transition.)
2. Calculates $D_{min.}$ D_{min} is the minimum computed cost to network *x*.
3. Determines membership of V_2. V_2 is the set of neighbors that are in V1 whose computed cost to network *x* equals D_{min}.

The feasibility condition is met when V_2 has one or more members.

The concept of feasible successors is illustrated in Figure 2-6. Consider Router A's topology table entries for Network 7. Router B is the *successor* with a computed cost of 31 to reach Network 7, compared to the computed costs of Router D (230) and Router H (40).

Figure 2-6 DUAL Feasible Successor



If Router B becomes unavailable, Router A will go through the following three-step process to find a feasible successor for Network 7:

Step 1 Determining which neighbors have an advertised distance to Network 7 that is less than Router A's feasible distance (FD) to Network 7. The FD is 31 and Router H meets this condition. Therefore, Router H is a member of V_1.

Step 2 Calculating the minimum computed cost to Network 7. Router H provides a cost of 40, and Router D provides a cost of 230. D_{min} is, therefore, 40.

Step 3 Determining the set of neighbors that are in V_1 whose computed cost to Network 7 equals D_{min} (40). Router H meets this condition.

The feasible successor is Router H which provides a least cost route of 40 from Router A to Network 7.

If Router H now also becomes unavailable, Router A performs the following computations:

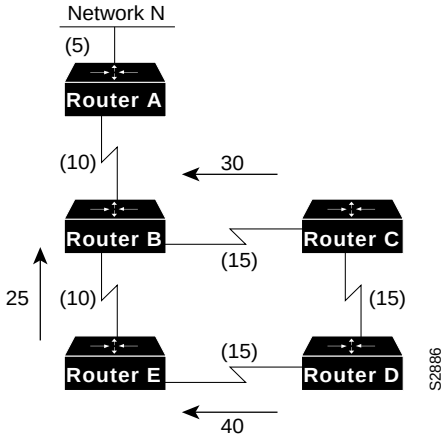
Step 1 Determines which neighbors have an advertised distance to Network 7 that is less than the FD for Network 7. Because both Router B and H have become unavailable, only Router D remains. However, the advertised cost of Router D to Network 7 is 220, which is greater than Router A's FD (31) to Network 7. Router D, therefore, cannot be a member of V_1. The FD remains at 31—the FD can only change during an active-to-passive transition, and this did not occur. There was no transition to active state for Network 7; this is known as a *local computation*.

Step 2 Because there are no members of V1, there can be no feasible successors. Router A, therefore, transitions from passive to active state for Network 7 and queries its neighbors about Network 7. There was a transition to active; this is known as a *diffusing computation.*

Note For more details on Enhanced IGRP convergence, see Appendix F, "References and Recommended Reading," for a list of reference papers and materials.

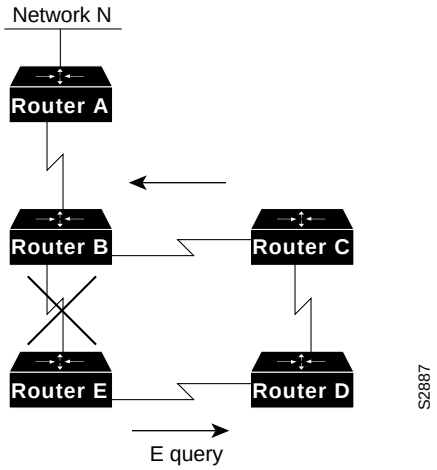
The following example and graphics further illustrate how Enhanced IGRP supports virtually instantaneous convergence in a changing internetwork environment. In Figure 2-7, all routers can access each other and Network N. The computed cost to reach other routers and Network N is shown. For example, the cost from Router E to Router B is 10. The cost from Router E to Network N is 25 (cumulative of 10 + 10 + 5 = 25).

Figure 2-7 DUAL Example (Part 1): Initial Network Connectivity



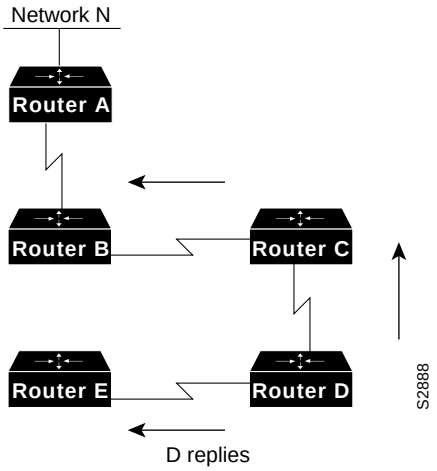
In Figure 2-8, the connection between Router B and Router E fails. Router E sends a multicast query to all of its neighbors and puts Network N into an active state.

Figure 2-8 DUAL Example (Part 2): Sending Queries



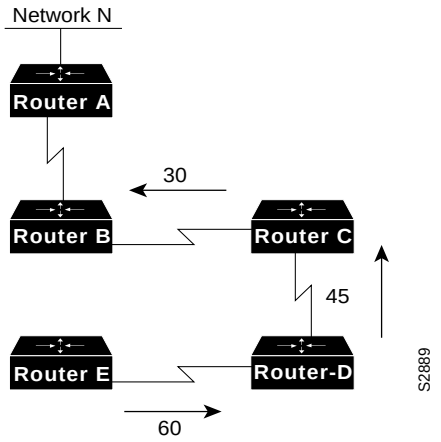
Next, as illustrated in Figure 2-9, Router D determines that it has a feasible successor. It changes its successor from Router E to Router C and sends a reply to Router E.

Figure 2-9 DUAL Example (Part 3): Switching to a Feasible Successor



In Figure 2-10, Router E has received replies from all neighbors and therefore brings Network N out of active state. Router E puts Network N into its routing table at a distance of 60.

Figure 2-10 DUAL Example (Part 4): Final Network Connectivity



Note Router A, Router B, and Router C were not involved in route recomputation. Router D recomputed its path to Network N without first needing to learn new routing information from its downstream neighbors.

Enhanced IGRP Network Scalability

Network scalability is limited by two factors: operational issues and technical issues. Operationally, Enhanced IGRP provides easy configuration and growth. Technically, Enhanced IGRP uses resources at less than a linear rate with the growth of a network.

Memory

A router running Enhanced IGRP stores all routes advertised by neighbors so that it can adapt quickly to alternate routes. The more neighbors a router has, the more memory a router uses. Enhanced IGRP automatic route aggregation bounds the routing table growth naturally. Additional bounding is possible with manual route aggregation.

CPU

Enhanced IGRP uses the DUAL algorithm to provide fast convergence. DUAL recomputes only routes which are affected by a topology change. DUAL is not computationally complex, so it does not require a lot of CPU.

Bandwidth

Enhanced IGRP uses partial updates. Partial updates are generated only when a change occurs; only the changed information is sent, and this changed information is sent only to the routers affected. Because of this, Enhanced IGRP is very efficient in its usage of bandwidth. Some additional bandwidth is used by Enhanced IGRP's HELLO protocol to maintain adjacencies between neighboring routers.

Enhanced IGRP Security

Enhanced IGRP is available only on Cisco routers. This prevents accidental or malicious routing disruption caused by hosts in a network.

In addition, route filters can be set up on any interface to prevent learning or propagating routing information inappropriately.

OSPF Internetwork Design Guidelines

OSPF is an Interior Gateway Protocol (IGP) developed for use in Internet Protocol (IP)-based internetworks. As an IGP, OSPF distributes routing information between routers belonging to a single autonomous system (AS). An AS is a group of routers exchanging routing information via a common routing protocol. The OSPF protocol is based on shortest-path-first, or link-state, technology.

The OSPF protocol was developed by the OSPF working group of the Internet Engineering Task Force (IETF). It was designed expressly for the Internet Protocol (IP) environment, including explicit support for IP subnetting and the tagging of externally derived routing information. OSPF Version 2 is documented in Request for Comments (RFC) 1247.

Whether you are building an OSPF internetwork from the ground up or converting your internetwork to OSPF, the following design guidelines provide a foundation from which you can construct a reliable, scalable OSPF-based environment.

Two design activities are critically important to a successful OSPF implementation:

- Definition of area boundaries
- Address assignment

Ensuring that these activities are properly planned and executed will make all the difference in your OSPF implementation. Each is addressed in more detail with the discussions that follow. These discussions are divided into six sections:

- OSPF Network Topology
- OSPF Addressing and Route Summarization
- OSPF Route Selection
- OSPF Convergence
- OSPF Network Scalability
- OSPF Security

Note For a detailed case study on how to set up and configure RIP and OSPF redistribution, see Chapter 1, “RIP and OSPF Redistribution” in the *Internetworking Case Studies* publication.

OSPF Network Topology

OSPF works best in a hierarchical routing environment. The first and most important decision when designing an OSPF network is to determine which routers and links are to be included in the backbone and which are to be included in each area.

There are several important guidelines to consider when designing an OSPF topology:

- **The number of routers in an area**—OSPF uses a CPU-intensive algorithm. The number of calculations that must be performed given *n* link-state packets is proportional to *n log n*. As a result, the larger and more unstable the area, the greater the likelihood for performance problems associated with routing protocol recalculation. Generally, an area should have no more than 50 routers. Areas with unstable links should be smaller.
- **The number of neighbors for any one router**—OSPF floods all link-state changes to all routers in an area. Routers with many neighbors have the most work to do when link-state changes occur. In general, any one router should have no more than 60 neighbors.
- **The number of areas supported by any one router**—A router must run the link-state algorithm for each link-state change that occurs for every area in which the router resides. Every area border router is in at least two areas (the backbone and one area). In general, to maximize stability, one router should not be in more than three areas.
- **Designated router selection**—In general, the designated router and backup designated router on a local-area network (LAN) have the most OSPF work to do. It is a good idea to select routers that are not already heavily loaded with CPU-intensive activities to be the designated router and backup designated router. In addition, it is generally not a good idea to select the same router to be designated router on many LANs simultaneously.

The discussions that follow address topology issues that are specifically related to the backbone and the areas.

Backbone Considerations

Stability and *redundancy* are the most important criteria for the backbone. Stability is increased by keeping the size of the backbone reasonable. This is caused by the fact that every router in the backbone needs to recompute its routes after every link-state change. Keeping the backbone small reduces the likelihood of a change and reduces the amount of CPU cycles required to recompute routes. As a general rule, each area (including the backbone) should contain no more than 50 routers. If link quality is high and the number of routes is small, the number of routers can be increased.

Redundancy is important in the backbone to prevent partition when a link fails. Good backbones are designed so that no single link failure can cause a partition.

OSPF backbones must be contiguous. All routers in the backbone should be directly connected to other backbone routers. OSPF includes the concept of virtual links. A virtual link creates a path between two area border routers (an area border router is a router connects an area to the backbone) that are not directly connected. A virtual link can be used to heal a partitioned backbone. However, it is not a good idea to design an OSPF network to require the use of virtual links. The stability of a virtual link is determined by the stability of the underlying area. This dependency can make troubleshooting more difficult. In addition, virtual links cannot run across stub areas. See the section "Backbone-to-Area Route Advertisement," later in this chapter for a detailed discussion of stub areas.

Avoid placing hosts (such as workstations, file servers or other shared resources) in the backbone area. Keeping hosts out of the backbone area simplifies internetwork expansion and creates a more stable environment.

Area Considerations

Individual areas must be contiguous. In this context, a contiguous area is one in which a continuous path can be traced from any router in an area to any other router in the same area. This does not mean that all routers must share a common network media. It is not possible to use virtual links to connect a partitioned area. Ideally, areas should be richly connected internally to prevent partitioning.

The two most critical aspects of area design follow:

- Determining how the area is addressed
- Determining how the area is connected to the backbone

Areas should have a contiguous set of network and/or subnet addresses. Without a contiguous address space, it is not possible to implement route summarization. The routers that connect an area to the backbone are called *area border routers*. Areas can have a single area border router or they can have multiple area border routers. In general, it is desirable to have more than one area border router per area to minimize the chance of the area becoming disconnected from the backbone.

When creating large-scale OSPF internetworks, the definition of areas and assignment of resources within areas must be done with a pragmatic view of your internetwork. The following are general rules that will help ensure that your internetwork remains flexible and provides the kind of performance needed to deliver reliable resource access.

- Consider physical proximity when defining areas—If a particular location is densely connected, create an area specifically for nodes at that location.
- Reduce the maximum size of areas if links are unstable—If your internetwork includes unstable links, consider implementing smaller areas to reduce the effects of route flapping. Whenever a route is lost or comes online, each affected area must converge on a new topology. The Dykstra algorithm will run on all the affected routers. By segmenting your internetwork into smaller areas, you can isolate unstable links and deliver more reliable overall service.

OSPF Addressing and Route Summarization

Address assignment and route summarization are inextricably linked when designing OSPF internetworks. To create a scalable OSPF internetwork, you should implement route summarization. To create an environment capable of supporting route summarization, you must implement an effective hierarchical addressing scheme. The addressing structure that you implement can have a profound impact on the performance and scalability of your OSPF internetwork. The following sections discuss OSPF route summarization and three addressing options:

- Separate network numbers for each area
- Network Information Center (NIC)-authorized address areas created using bit-wise subnetting and VLSM
- Private addressing, with a “demilitarized zone” (DMZ) buffer to the official Internet world

---

Note You should keep your addressing scheme as simple as possible, but be wary of oversimplifying your address assignment scheme. Although simplicity in addressing saves time later when operating and troubleshooting your network, taking short cuts can have certain severe consequences. In building a scalable addressing environment, use a structured approach. If necessary, use bit-wise subnetting—but make sure that route summarization can be accomplished at the area border routers.

OSPF Route Summarization

Route summarization is extremely desirable for a reliable and scalable OSPF internetwork. The effectiveness of route summarization, and your OSPF implementation in general, hinges on the addressing scheme that you adopt. Summarization in an OSPF internetwork occurs between each area and the backbone area. Summarization must be configured manually in OSPF.

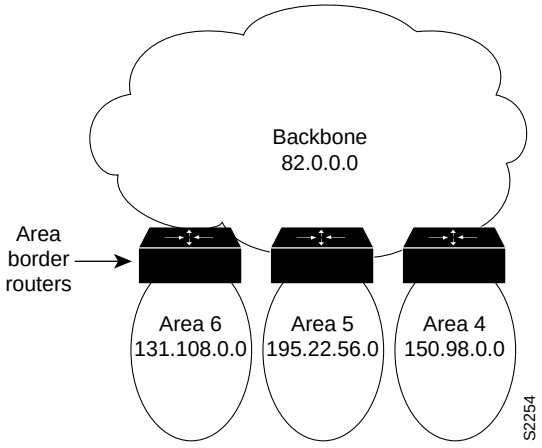
When planning your OSPF internetwork, consider the following issues:

- Be sure that your network addressing scheme is configured so that the range of subnets assigned within an area is contiguous.
- Create an address space that will permit you to split areas easily as your network grows. If possible, assign subnets according to simple octet boundaries. If you cannot assign addresses in an easy-to-remember and easy-to-divide manner, be sure to have a thoroughly defined addressing structure. If you know how your entire address space is assigned (or will be assigned), you can plan for changes more effectively.
- Plan ahead for the addition of new routers to your OSPF environment. Be sure that new routers are inserted appropriately as area, backbone, or border routers. Because the addition of new routers creates a new topology, inserting new routers can cause unexpected routing changes (and possibly performance changes) when your OSPF topology is recomputed.

Separate Address Structures for Each Area

One of the simplest ways to allocate addresses in OSPF is to assign a separate network number for each area. With this scheme, you create a backbone and multiple areas, and assign a separate IP network number to each area. Figure 2-11 illustrates this kind of area allocation.

Figure 2-11 Assignment of NIC Addresses Example



The following are the basic steps for creating such a network:

Step 1 Define your structure (identify areas and allocate nodes to areas).

Step 2 Assign addresses to networks, subnets, and end stations.

In the network illustrated in Figure 2-11, each area has its own unique NIC-assigned address. These can be Class A (the backbone in Figure 2-11), Class B (areas 4 and 6), or Class C (Area 5).

The following are some clear benefits of assigning separate address structures to each area:

- Address assignment is relatively easy to remember.
- Configuration of routers is relatively easy and mistakes are less likely.
- Network operations are streamlined because each area has a simple, unique network number.

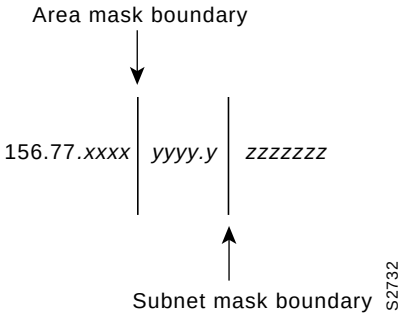
In the example illustrated in Figure 2-11, the route summarization configuration at the area border routers is greatly simplified. Routes from Area 4 injecting into the backbone can be summarized as follows: *all routes starting with 150.98 are found in Area 4*.

The main drawback of this approach to address assignment is that it wastes address space. If you decide to adopt this approach, be sure that area border routers are configured to do route summarization. Summarization must be explicitly set; it is disabled by default in OSPF.

Bit-Wise Subnetting and VLSM

Bit-wise subnetting and variable-length subnetwork masks (VLSMs) can be used in combination to save address space. Consider a hypothetical network where a Class B address is subdivided using an area mask and distributed among 16 areas. The Class B network, 156.77.0.0, might be subdivided as illustrated in Figure 2-12.

Figure 2-12 Areas and Subnet Masking



In Figure 2-12, the letters *x*, *y*, and *z* represent bits of the last two octets of the Class B network as follows:

- The four *x* bits are used to identify 16 areas.
- The five *y* bits represent up to 32 subnets per area.
- The seven *z* bits allow for 126 (128-2) hosts per subnet.

Appendix A, “Subnetting an IP Address Space” provides a complete example illustrating assignment for the Class B address 150.100.0.0. It illustrates both the concept of *area masks* and the breakdown of large subnets into smaller ones using VLSMs.

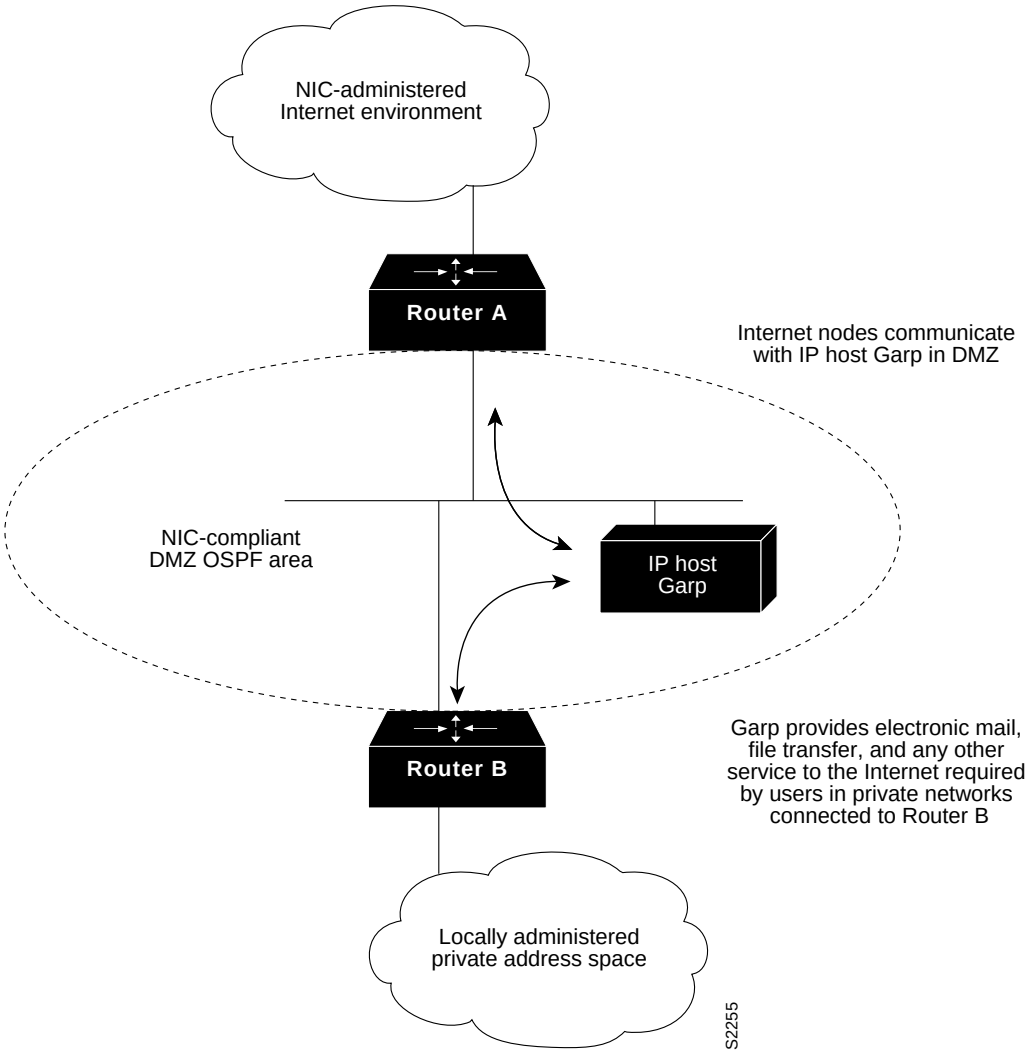
Private Addressing

Private addressing is another option often cited as simpler than developing an area scheme using bit-wise subnetting. Although private address schemes provide an excellent level of flexibility and do not limit the growth of your OSPF internetwork, they have certain disadvantages. For instance, developing a large-scale internetwork of privately addressed IP nodes limits total access to the

Internet, and mandates the implementation of what is referred to as a *demilitarized zone* (DMZ). If you need to connect to the Internet, Figure 2-13 illustrates the way in which a DMZ provides a buffer of valid NIC nodes between a privately addressed network and the Internet.

All nodes (end systems and routers) on the network in the DMZ must have NIC-assigned IP addresses. The NIC might, for example, assign a single Class C network number to you. The DMZ shown in Figure 2-13 has two routers and a single application gateway host (Garp). Router A provides the interface between the DMZ and the Internet, and Router B provides the firewall between the DMZ and the private address environment. All applications that need to run over the Internet must access the Internet through the application gateway.

Figure 2-13 Connecting to the Internet from a Privately Addressed Network



Note For a case study on network security that includes information on how to set up firewall routers and communication servers, see Chapter 3, "Increasing Security on IP Networks" in the *Internetworking Case Studies* publication.

Route Summarization Techniques

Route summarization is particularly important in an OSPF environment because it increases the stability of the network. If route summarization is being used, routes within an area that change do not need to be changed in the backbone or in other areas.

Route summarization addresses two important questions of route information distribution:

- What information does the backbone need to know about each area? The answer to this question focuses attention on area-to-backbone routing information.
- What information does each area need to know about the backbone and other areas? The answer to this question focuses attention on backbone-to-area routing information.

Area-to-Backbone Route Advertisement

There are several key considerations when setting up your OSPF areas for proper summarization:

- OSPF route summarization occurs in the area border routers.
- OSPF supports VLSM, so it is possible to summarize on any bit boundary in a network or subnet address.
- OSPF requires manual summarization. As you design the areas, you need to determine summarization at each area border router.

Backbone-to-Area Route Advertisement

There are four potential types of routing information in an area:

- Default. If an explicit route cannot be found for a given IP network or subnetwork, the router will forward the packet to the destination specified in the default route.
- Intra-area routes. Explicit network or subnet routes must be carried for all networks or subnets inside an area.
- Interarea routes. Areas may carry explicit network or subnet routes for networks or subnets that are in this AS but not in this area.
- External routes. When different ASs exchange routing information, the routes they exchange are referred to as external routes.

In general, it is desirable to restrict routing information in any area to the minimal set that the area needs.

There are three types of areas, and they are defined in accordance with the routing information that is used in them:

- Non-stub areas—Non-stub areas carry a default route, static routes, intra-area routes, interarea routes and external routes. An area must be a nonstub area when it contains a router that uses both OSPF and any other protocol, such as the Routing Information Protocol (RIP). Such a router is known as an autonomous system border router (ASBR). An area must also be a nonstub area when a virtual link is configured across the area. Nonstub areas are the most resource-intensive type of area.
- Stub areas—Stub areas carry a default route, intra-area routes and interarea routes, but they do not carry external routes. Stub areas are recommended for areas that have only one area border router and they are often useful in areas with multiple area border routers. See “Controlling Interarea Traffic,” later in this chapter for a detailed discussion of the design trade-offs in areas with multiple area border routers.There are two restrictions on the use of stub areas: virtual links cannot be configured across them, and they cannot contain an ASBR.

- Stub areas without summaries—Software releases 9.1(11), 9.21(2), and 10.0(1) and later support stub areas without summaries, allowing you to create areas that carry only a default route and intra-area routes. Stub areas without summaries do not carry interarea routes or external routes. This type of area is recommended for simple configurations where a single router connects an area to the backbone.

Table 2-2 shows the different types of areas according to the routing information that they use.

Table 2-2 Routing Information Used in OSPF Areas

Area Type	Default Route	Intra-area Routes	Interarea Routes	External Routes
Nonstub	Yes	Yes	Yes	Yes
Stub	Yes	Yes	Yes	No
Stub without summaries	Yes	Yes	No	No

Note Stub areas are configured using the **area** *area-id* **stub** router configuration command. Routes are summarized using the **area** *area-id* **range** *address mask* router configuration command. Refer to your *Router Products Configuration Guide* and *Router Products Command Reference* publications for more information regarding the use of these commands.

OSPF Route Selection

When designing an OSPF internetwork for efficient route selection, consider three important topics:

- Tuning OSPF Metrics
- Controlling Interarea Traffic
- Load Balancing in OSPF Internetworks

Tuning OSPF Metrics

The default value for OSPF metrics is based on bandwidth. The following characteristics show how OSPF metrics are generated:

- Each link is given a metric value based on its bandwidth. The metric for a specific link is the inverse of the bandwidth for that link. Link metrics are normalized to give FDDI a metric of 1. The metric for a route is the sum of the metrics for all the links in the route.

Note In some cases, your network might implement a media type that is faster than the fastest default media configurable for OSPF (FDDI). An example of a faster media is ATM. By default, a faster media will be assigned a cost equal to the cost of an FDDI link—a link-state metric cost of 1. Given an environment with both FDDI and a faster media type, you must manually configure link costs to configure the faster link with a lower metric. Configure any FDDI link with a cost greater than 1, and the faster link with a cost less than the assigned FDDI link cost. Use the **ip ospf cost** interface configuration command to modify link-state cost.

- When route summarization is enabled, OSPF uses the metric of the best route in the summary.

- There are two forms of external metrics: type 1 and type 2. Using an external type 1 metric results in routes adding the internal OSPF metric to the external route metric. External type 2 metrics do not add the internal metric to external routes. The external type 1 metric is generally preferred. If you have more than one external connection, either metric can affect how multiple paths are used.

Controlling Interarea Traffic

When an area has only a single area border router, all traffic that does not belong in the area will be sent to the area border router.

In areas that have multiple area border routers, two choices are available for traffic that needs to leave the area:

- Use the area border router closest to the originator of the traffic. (Traffic leaves the area as soon as possible.)
- Use the area border router closest to the destination of the traffic. (Traffic leaves the area as late as possible.)

If the area border routers inject only the default route, the traffic goes to the area border router that is closest to the source of the traffic. Generally, this behavior is desirable because the backbone typically has higher bandwidth lines available. However, if you want the traffic to use the area border router that is nearest the destination (so that traffic leaves the area as late as possible), the area border routers should inject summaries into the area instead of just injecting the default route.

Most network designers prefer to avoid asymmetric routing (that is, using a different path for packets that are going from A to B than for those packets that are going from B to A.) It is important to understand how routing occurs between areas to avoid asymmetric routing.

Load Balancing in OSPF Internetworks

Internetwork topologies are typically designed to provide redundant routes in order to prevent a partitioned network. Redundancy is also useful to provide additional bandwidth for high traffic areas. If equal-cost paths between nodes exist, Cisco routers automatically load balance in an OSPF environment.

Cisco routers can use up to four equal-cost paths for a given destination. Packets might be distributed either on a per-destination (when fast switching) or a per-packet basis. Per-destination load balancing is the default behavior. Per-packet load balancing can be enabled by turning off fast switching using the **no ip route-cache** interface configuration command. For line speeds of 56 kbps and faster, it is recommended that you enable fast switching.

OSPF Convergence

One of the most attractive features about OSPF is the ability to quickly adapt to topology changes.

There are two components to routing convergence:

- Detection of topology changes—OSPF uses two mechanisms to detect topology changes. Interface status changes (such as carrier failure on a serial link) is the first mechanism. The second mechanism is failure of OSPF to receive a hello packet from its neighbor within a timing window called a *dead timer.* Once this timer expires, the router assumes the neighbor is down. The dead timer is configured using the **ip ospf dead-interval** interface configuration command. The default value of the dead timer is four times the value of the Hello interval. That results in a dead timer default of 40 seconds for broadcast networks and 2 minutes for nonbroadcast networks.

- Recalculation of routes—Once a failure has been detected, the router that detected the failure sends a link-state packet with the change information to all routers in the area. All the routers recalculate all of their routes using the Dykstra (or SPF) algorithm. The time required to run the algorithm depends on a combination of the size of the area and the number of routes in the database.

OSPF Network Scalability

Your ability to scale an OSPF internetwork depends on your overall network structure and addressing scheme. As outlined in the preceding discussions concerning network topology and route summarization, adopting a hierarchical addressing environment and a structured address assignment will be the most important factors in determining the scalability of your internetwork.

Network scalability is affected by operational and technical considerations:

- Operationally, OSPF networks should be designed so that areas do not need to be split to accommodate growth. Address space should be reserved to permit the addition of new areas.
- Technically, scaling is determined by the utilization of three resources: memory, CPU, and bandwidth.

Memory

An OSPF router stores all of the link states for all of the areas that it is in. In addition, it can store summaries and externals. Careful use of summarization and stub areas can reduce memory use substantially.

CPU

An OSPF router uses CPU cycles whenever a link-state change occurs. Keeping areas small and using summarization dramatically reduces CPU use and creates a more stable environment for OSPF.

Bandwidth

OSPF sends partial updates when a link-state change occurs. The updates are flooded to all routers in the area. In a quiet network, OSPF is a quiet protocol. In a network with substantial topology changes, OSPF minimizes the amount of bandwidth used.

OSPF Security

Two kinds of security are applicable to routing protocols:

- Controlling the routers that participate in an OSPF network

  OSPF contains an optional authentication field. All routers within an area must agree on the value of the authentication field. Because OSPF is a standard protocol available on many platforms, including some hosts, using the authentication field prevents the inadvertent startup of OSPF in an uncontrolled platform on your network and reduces the potential for instability.

- Controlling the routing information that routers exchange

All routers must have the same data within an OSPF area. As a result, it is not possible to use route filters in an OSPF network to provide security.