CHAPTER 1

Internetworking Design Basics

Designing an internetwork can be a challenging task. An internetwork that consists of only 50 meshed routing nodes can pose complex problems that lead to unpredictable results. Attempting to optimize internetworks that feature thousands of nodes can pose even more complex problems.

Despite improvements in equipment performance and media capabilities, internetwork design is becoming more difficult. The trend is toward increasingly complex environments involving multiple media, multiple protocols, and interconnection to networks outside any single organization's dominion of control. Carefully designing internetworks can reduce the hardships associated with growth as a networking environment evolves.

This chapter provides an overview of planning and design guidelines. Discussions are divided into the following general topics:

- Determining Your Internetworking Requirements
- Identifying and Selecting Internetworking Capabilities
- Choosing Internetworking Reliability Options

Determining Your Internetworking Requirements

Routers and other internetworking devices must reflect the goals, characteristics, and policies of the organizations in which they operate.

Two primary goals drive internetworking design and implementation:

- Application availability—Networks carry application information between computers. If the applications are not available to network users, the network is not doing its job.
- Cost of ownership—Information system (IS) budgets today often run in the millions of dollars. As large organizations increasingly rely on electronic data for managing business activities, the associated costs of computing resources will continue to rise.

A well-designed internetwork can help to balance these objectives. When properly implemented, routers can optimize application availability while allowing for the cost-effective use of existing network resources.

The Design Problem: Optimizing Availability and Cost

In general, the network design problem consists of three general elements:

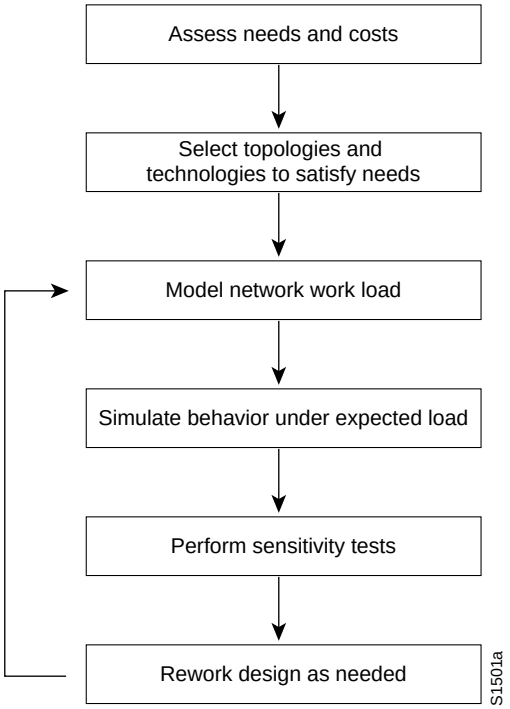
- *Environmental givens*—Environmental givens include the location of hosts, servers, terminals and other end nodes; the projected traffic for the environment; and the projected costs for delivering different service levels.
- *Performance constraints*—Performance constraints consist of network reliability and traffic throughput.
- *Internetworking variables*— Internetworking variables include the network topology, line capacities, and packet flow assignments.

The goal is to minimize cost based on these elements while delivering service that does not compromise established availability requirements.

Internetwork designers face two primary concern— availability and cost. These issues are essentially at odds. Any increase in availability must generally be reflected as an increase in cost. As a result, network designers must weigh the relative importance of resource availability and overall cost carefully.

Designing your network is an iterative activity. Figure 1-1 illustrates the basic process associated with internetwork design. The discussions that follow outline several areas that require careful consideration when planning your internetworking implementation.

Figure 1-1 General Network Design Process



Assessing User Requirements

In general, users primarily want application availability in their networks. The chief components of application availability are *response time*, *throughput*, and *reliability.*

Response time is the time between entry of a command or keystroke and the host system's execution of the command, or delivery of a response. User satisfaction about response time is generally considered to be a *monotonic* function up to some limit, at which point user satisfaction falls off to nearly zero. Applications where fast response time is considered critical include interactive online services such as automated tellers and point-of-sale machines.

Applications that put high-volume traffic onto the network have more effect on throughput than end-to-end connections. Throughput-intensive applications generally involve file-transfer activities. However, throughput-intensive applications also usually have low response-time requirements. Indeed, they can often be scheduled at times when response-time-sensitive traffic is low (for example, after normal work hours).

Although reliability is always important, some applications have genuine requirements that exceed typical needs. Organizations that require nearly 100 percent up time conduct all activities online or over the telephone. Financial services, securities exchanges, and emergency/police/military operations are a few examples. These situations imply a requirement for a high level of hardware and topological redundancy. Determining the cost of any downtime is essential in determining the relative importance of reliability to your internetwork.

You can assess user requirements in a number of ways. The more involved your users are in the process, the more likely that your evaluation will be accurate. In general, you can use the following methods to get this information:

- **User community profiles**—Outline what different user groups require. This is the first step in determining internetwork requirements. Although many users have roughly the same requirements of an electronic mail system, engineering groups using X Windows terminals and Sun workstations in an NFS environment have different needs from PC users sharing print servers in a finance department.
- **Interviews, focus groups, and surveys**—Build a baseline for implementing an internetwork. Understand that some groups might require access to common servers. Others might want to allow external access to specific internal computing resources. Certain organizations might require IS support systems to be managed in a particular way according to some external standard. The least formal method of obtaining information is to conduct interviews with key user groups. Focus groups can also be used to gather information and generate discussion among different organizations with similar (or dissimilar) interests. Finally, formal surveys can be used to get a statistically valid reading of user sentiment regarding a particular service level or proposed internetworking architecture.
- **Human factors tests**—The most expensive, time-consuming, and possibly revealing method is to conduct a test involving representative users in a lab environment. This is most applicable when evaluating response time requirements. As an example, you might set up working systems and have users perform normal remote host activities from the lab network. By evaluating user reactions to variations in host responsiveness, you can create benchmark thresholds for acceptable performance.

Compatibility, Conformance, and Interoperability

Compatibility, conformance, and interoperability are related to the problem of balancing proprietary functionality and open internetworking flexibility. As a network designer, you might be forced to choose between implementing a multivendor environment and implementing a specific, proprietary capability.

For example, the Interior Gateway Routing Protocol (IGRP) provides many useful capabilities, such as fast convergence and efficient route handling in large internetworks, but it is a proprietary routing protocol. In contrast, the integrated Intermediate System-to-Intermediate System (IS-IS) protocol is an open internetworking alternative that also provides a fast converging routing environment; however, implementing an open routing protocol can potentially result in greater multivendor configuration complexity.

The protocol that you choose will have far-ranging effects on your overall internetwork design. Assume that you decide to implement integrated IS-IS instead of IGRP. In doing this, you gain a measure of interoperability; however, you lose some functionality. For instance, you will not be able to load balance traffic over unequal parallel paths. Similarly, some modems provide a high level of proprietary diagnostic capabilities, but require that all modems throughout a network be of the same vendor type to fully exploit proprietary diagnostics.

Previous internetworking (and networking) investments and expectations for future requirements have considerable influence over your choice of implementations. You need to consider installed internetworking and networking equipment; applications running (or to be run) on the network; physical location of sites, hosts, and users; rate or growth of user community; and both physical and logical network layout.

Assessing Costs

The internetwork is a strategic element in your overall information system design. As such, the cost of your internetwork is much more than the sum of your router purchase orders. View it as a total cost-of-ownership issue. You must consider the entire life cycle of your internetworking environment. A brief list of costs associated with internetworks follows:

- **Router hardware and software costs**—Consider what is really being bought when you purchase your systems; costs should include initial purchase and installation, maintenance, and projected upgrade costs.
- **Performance tradeoff costs**—Consider the cost of going from a 5-second response time to a half-second response time. Such improvements can cost quite a bit in terms of media selection, network interfaces, internetworking nodes, modems, and wide-area network (WAN) services.
- **Installation costs**—Installing a site's physical cable plant is by far the most expensive element of a large network. The costs include installation labor, site modification, fees associated with local code conformance, and costs incurred to ensure compliance with environmental restrictions (such as asbestos removal). Other important elements in keeping your costs to a minimum will include developing a well-planned wiring closet layout and implementing color code conventions for cable runs.
- **Expansion costs**—Calculate the cost of ripping out all thick Ethernet, adding additional functionality, or moving to a new location. Projecting your future requirements and accounting for future needs saves time and money.
- **Support costs**—Complicated internetworks cost more to monitor, configure, and maintain. Your internetwork should be no more complicated than necessary. Costs include training, direct labor (network managers and administrators), sparing, and replacement costs.
- **Cost of downtime**—Evaluate the cost for every minute that a user is unable to access a file server or a centralized database. If this cost is high, you must attribute a high cost to downtime. If the cost is high enough, fully redundant internetworks might be your only option.
- **Opportunity costs**—Every choice you make will have an opposing alternative option. Whether that option is a specific hardware platform, topology solution, level of redundancy, or system integration alternative, there are always options. *Opportunity costs* are the costs of not picking one of those options. The opportunity costs of not switching to newer technologies and

topologies might be lost competitive advantage, lower productivity, and slower overall performance. Any effort to integrate opportunity costs into your analysis can help to make accurate comparisons at the beginning of your project.

- **Sunken costs**—Your investment in existing cable plant, routers, concentrators, hosts, and other equipment and software are your *sunken costs*. If the sunken cost is high, you might need to modify your networks so that your existing internetwork can continue to be exploited. Although comparatively low incremental costs might appear to be more attractive than significant redesign costs, your organization might pay more in the long run by not upgrading systems. Over-reliance on sunken costs can cost your organization sales and market share when calculating the cost of internetwork modifications and additions.

Estimating Traffic: Work Load Modeling

Empirical *work-load modeling* consists of instrumenting a working internetwork and monitoring traffic for a given number of users, applications, and network topology. Try to characterize activity throughout a normal work day in terms of the type of traffic passed, level of traffic, response time of hosts, time to execute file transfers, and so on. You can also observe utilization on existing routers over the test period with the router's **show** and **debug** commands.

If the tested internetwork's characteristics are close to the new internetwork, you can try extrapolating to the new internetwork's number of users, applications, and topology. This is a *best-guess* approach to traffic estimation given the unavailability of tools to characterize this behavior in a dynamically routed environment.

In addition to passive monitoring of an existing network, you can measure activity and traffic generated by a known number of users attached to a representative test network and then extrapolate findings to your anticipated population.

One problem with modeling work loads on networks is that it is difficult to accurately pinpoint traffic load and network device performance as functions of the number of users, type of application, and geographical location. This is especially true without a real network in place. Consider the following factors that influence the dynamics of the network:

- The time-dependent nature of network access—Peak periods can vary; measurements must reflect a range of observations that includes peak demand.
- Differences associated with type of traffic—Routed and bridged traffic place different demands on internetwork devices and protocols; some protocols are sensitive to dropped packets; some application types require more bandwidth.
- The random (nondeterministic) nature of network traffic—Exact arrival time and specific effects of traffic are unpredictable.
- Competing protocols—In multiprotocol environments, routers undergo dynamic demands from different users, resulting in unpredictable effects.

Sensitivity Testing

From a practical point of view, sensitivity testing involves breaking stable links and observing what happens. When working with a test network, this is relatively easy. Perturb the network by removing an active interface and monitor how the change is handled by the internetwork: how traffic is rerouted, the speed of convergence, whether any connectivity is lost, and whether problems arise in handling specific types of traffic. You can also change the level of traffic on a network to determine the effects on the network when traffic levels approach media saturation. This empirical testing is a

type of *regression* testing: a series of specific modifications (tests) are repeated on different versions of network configurations. By monitoring the effects on the design variations, you can characterize the relative resilience of the design.

Modeling sensitivity tests using a computer is beyond the scope of this publication. A useful source for more information about computer-based network design and simulation is Andrew S.Tannenbaum's book, *Computer Networks*.

Identifying and Selecting Internetworking Capabilities

Once you understand your internetworking requirements, you must identify and then select the specific capabilities that fit your computing environment. The following discussions provide a starting point for making these decisions. Three topics are addressed:

- Contrasting Bridging and Routing Capabilities
- General Hierarchical Model for Internetworking
- Capabilities Associated with Backbone, Distribution, and Local-Access Services

---

Note This material introduces the primary router capabilities and outlines how each fulfills various internetworking requirements. The technology chapters that follow, addressing routing protocol implementation, IBM internetworking, and packet-service internetworking, present detailed discussions about specific implementations of large-scale internetworking.

Contrasting Bridging and Routing Capabilities

Data communications experts generally agree that bridges and routers are moving away from once-clear distinctions between the two technologies and converging toward the all-in-one brouter, routing bridge, or router/bridge. Performance enhancements are making the question of which is better (bridges or routers) irrelevant. The discussion that follows outlines the key criteria to use when determining which technology best suits your situation.

Bridging and Routing Definitions

Before bridging and routing capabilities can be contrasted, you must understand where each falls within a common framework of internetworking terminology. For this comparison, the bridging discussion focuses on transparent bridging. Transparent bridging is emphasized here because source-route bridging (SRB) has more in common with routing than with bridging. Refer to the *Internetworking Technology Overview* for more information about each of these technologies.

By convention, bridging is said to occur at the data-link layer, while routing is said to occur at the network layer of the International Organization for Standardization (ISO) seven-layer protocol model. In general, data-link layer devices assume a common logical network (information traverses a single hop to reach a destination). Network layer devices are designed to handle multiple hops and multiple networks. These distinctions lead to certain constraints for bridging that result in four important differences between routers and bridges:

- The header associated with data-link packets lacks information fields that are present in network layer packets. Examples of fields provided in network layer packets include final destination address, hop count, and fragmentation and reassembly information.

- Bridges do not support handshaking protocols, such as the Internet Control Message Protocol (ICMP) associated with the Internet Protocol (IP), used by end nodes and routers to learn about each other.
- Bridges cannot reorder packets from the same source; network layer protocols expect some degree of reordering (caused by fragmentation).
- Bridges use the Media Access Control (MAC) addresses defined at the time of equipment manufacture to identify an end node. Thus, the address has no topological meaning. With routers, a network address is associated with the local-area network (LAN) to which a particular node is attached.

Despite the constraints associated with bridging, there are often situations that require bridging technology. Similarly, routing might be necessary to ensure proper segmentation of traffic or to support a specific topology that does not permit a single logical internetwork. The following sections summarize common reasons for choosing routing or bridging.

Routing Advantages

Routing offers the following advantages over bridging:

- Routers can choose the best path that exists between source and destination; bridges are limited to a specific path (referred to as a spanning tree) through an internetwork. Two characteristics of bridges result in this difference: bridges must learn the location of stations based on the direction from which traffic is received, and bridges are transparent (in other words, they are not permitted to modify a packet in any way). These characteristics do not apply to source-route bridges. Unlike transparent bridges, source-route bridges do not maintain a station location table. In addition, loops are not possible in an SRB environment. The SRB standard states that an SRB must verify that the output network segment was not traversed before transmitting packets.
- Routers reconfigure topology after changes much more quickly than bridges, resulting in reduced service loss. Loops pose a substantially greater risk to a bridged internetwork than to a routed internetwork because bridges are transparent. Thus, new bridge paths are recognized gradually, while new routing paths are recognized as soon as routing information is received.
- The total number of stations that can be supported in a routed internetwork is virtually unlimited, particularly for ISO Connectionless Network Service (CLNS) internetworks, but the maximum number of stations in a bridged internetwork is constrained to thousands of end stations. Routers can accommodate a much larger address space because network layer addresses include information that groups nodes into areas or domains. Bridges have no hierarchical addressing component and cannot direct traffic in a hierarchical manner.
- Routers can provide a barrier against broadcast storms; bridges cannot. By design, when bridges join a series of LAN segments, those segments form a single LAN from the perspective of upper-layer protocols. A broadcast storm disables the entire bridged internetwork because all the bridges forward the broadcast traffic throughout the internetwork and are unable to intervene. Routers block broadcasts by default.
- Routers fragment and reassemble large packets; bridges drop packets that are too big to forward. For some protocols, the network layer header includes fragmentation and reassembly information; the data-link layer header does not. As a result, bridges simply drop packets that are too large to forward. In addition, bridges are unable to inform the source that the packet was dropped.
- Routers provide congestion feedback to end stations when traffic is heavy; bridges do not. In ISO CLNS, mechanisms in the network layer protocols provide congestion-based information that can be relayed to source nodes, forcing them to reduce transmission rates. The data-link layer provides no analogous capability.

Note In general, the preceding discussion concerning routing advantages assumes a comparison with transparent bridging. These advantages do no apply to SRB. However, one weakness of SRB is that it does not provide any form of dynamic rerouting. This is left to the end stations.

Bridging Advantages

Consider the following issues when choosing between routing or bridging:

- Bridges require minimal configuration; routers require maximum configuration. Some configuration is always required for routers. For example, IP routers require the configuration of separate addresses for each interface and substantial configuration of end nodes (addresses and masks). In some situations, basic learning bridges require virtually no configuration. To become operational, you can simply take the bridge out of the box, power it up, and attach it to a network.
- Bridges have a better price-to-performance ratio than routers. With less overhead to handle, bridges have enjoyed an advantage over routers in terms of pure packet traffic handling. However, because router performance has improved significantly (now providing near wire speed performance) and the price difference between routers and bridges has eroded, this advantage is diminishing.
- Bridges are protocol independent; routers are protocol dependent. As routers have become capable of handling multiple protocols as either *integrated* or *ships-in-the-night* environments (both of which are discussed in the "Backbone Routing Options" section later in this chapter), this perceived advantage is also diminishing. Nonetheless, bridges can handle multiple protocols with almost no configuration.
- Bridges forward nonroutable protocols such as Digital's Local Area Transport (LAT); routers cannot. Some protocols are not routable, even through they were designed to provide upper-layer functionality (just not between LANs). This remains a compelling reason for implementing bridging capabilities for supporting certain end-to-end connectivity.

Integrated Solutions

The trend in internetworking is to provide network designers greater flexibility in solving multiple internetworking problems without creating multiple networks or writing off existing data communications investments.

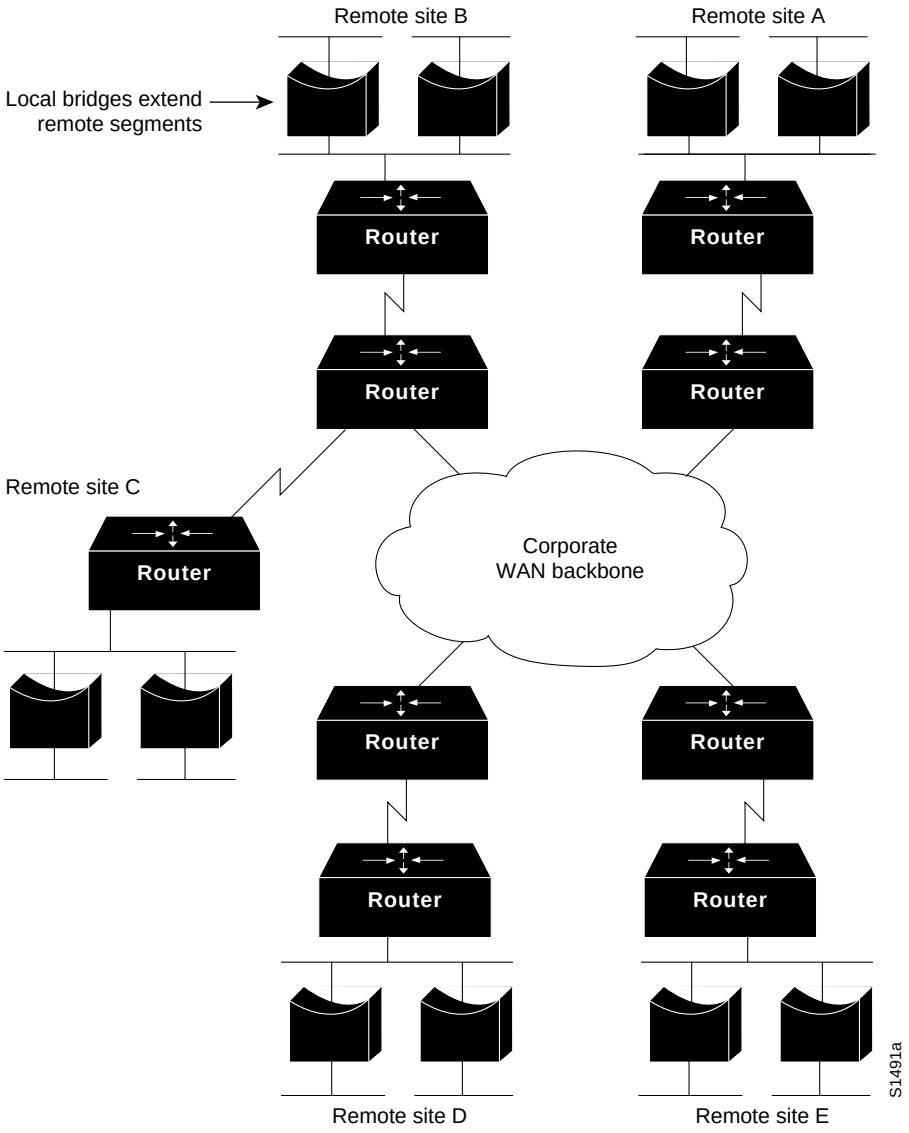
A network designer might employ bridges in a remote site for their ease of implementation, simple topology, and low traffic requirements. Routers might be relied upon to provide a reliable, self-healing backbone, as well as a barrier against inadvertent broadcast storms in the local networks.

A dedicated bridging implementation for simple internetworks (often remote sites) can be used in a number of circumstances. For example, an IGS can be implemented for local bridging at a remote site, while providing the option of routing as the remote network evolves. If you are interconnecting the remote site into a corporate backbone, it is best to implement routing at the point of access from the local internetwork to the backbone.

If you have a very large, meshed local internetwork (a campus consisting of several buildings and running protocols that derive significant benefits from routing), routers provide superior segmentation and more efficient traffic handling than bridges.

Figure 1-2 illustrates an environment that features a WAN backbone interconnecting remote sites to a corporate internetwork. Think of bridges that are implemented in this example as very smart repeaters.

Figure 1-2 Hybrid Internetwork Featuring Bridging and Routing Nodes



Backbone Routing Options

In an ideal world, the perfect enterprise-wide internetwork would feature a single, bullet-proof network protocol capable of transporting all manner of data communications seamlessly, error free, and with sufficient resilience to accommodate any unforeseen connectivity disruption. However, in the real world, there are many protocols with varying levels of resilience.

In designing a backbone for your organization, you might consider several options. These options are typically split into two primary categories:

- Multiprotocol Routing Backbone
- Single-Protocol Backbone

The following discussions outline the characteristics and properties of these two strategies.

Multiprotocol Routing Backbone

When multiple network layer protocols are routed throughout a common backbone without encapsulation (also referred to as *native* mode routing), the environment is referred to as a multiprotocol routing backbone. A multiprotocol backbone environment can adopt one of two routing strategies, or both, depending on the routed protocol involved. The two strategies are generally referred to as *integrated routing* and *ships in the night*.

Integrated routing involves the use of a single routing protocol (for example, a link state protocol) that determines the least cost path for different routed protocols.

The ships-in-the-night approach involves the use of a different routing protocol for each network protocol. For instance, some large-scale networks might feature multiple protocols where Novell IPX traffic is routed using a proprietary version of the Routing Information Protocol (RIP), IP is routed with IGRP, and DECnet Phase V traffic is routed via ISO CLNS-compliant IS-IS. Each of these network layer protocols is routed independently, with separate routing processes handling their traffic and separate paths calculated.

Mixing routers within an internetwork that supports different combinations of multiple protocols can create a confusing situation, particularly for integrated routing. In general, integrated routing is easier to manage if all the routers attached to the integrated routing backbone support the same integrated routing scheme. Routes for other protocols can be calculated separately. As an alternative, you can use encapsulation to transmit traffic over routers that do not support a particular protocol.

Single-Protocol Backbone

With a single-protocol backbone, all routers are assumed to support a single routing protocol for a single network protocol. In this kind of routing environment, all other routing protocols are ignored. If multiple protocols are to be passed over the internetwork, unsupported protocols must be encapsulated within the supported protocol or they will be ignored by the routing nodes.

Why implement a single-protocol backbone? If relatively few other protocols are supported at a limited number of isolated locations, it is reasonable to implement a single protocol backbone. However, encapsulation does add overhead to traffic on the network. If multiple protocols are supported widely throughout a large internetwork, a multiprotocol backbone approach is likely to work better.

In general, you should support all the network layer protocols in an internetwork with a native routing solution and implement as few network layer protocols as possible.

Hierarchical Internetworking Model

Most internetworks can be hierarchically divided into three logical services: backbone, distribution, and local-access internetworks. *Backbone* (or *core*) services aim to optimize communication among routers at different sites or in different logical groupings. *Distribution* services provide a way to implement policy-based traffic control to isolate backbone and local environments. *Local-access* services support communication between end stations and routers.

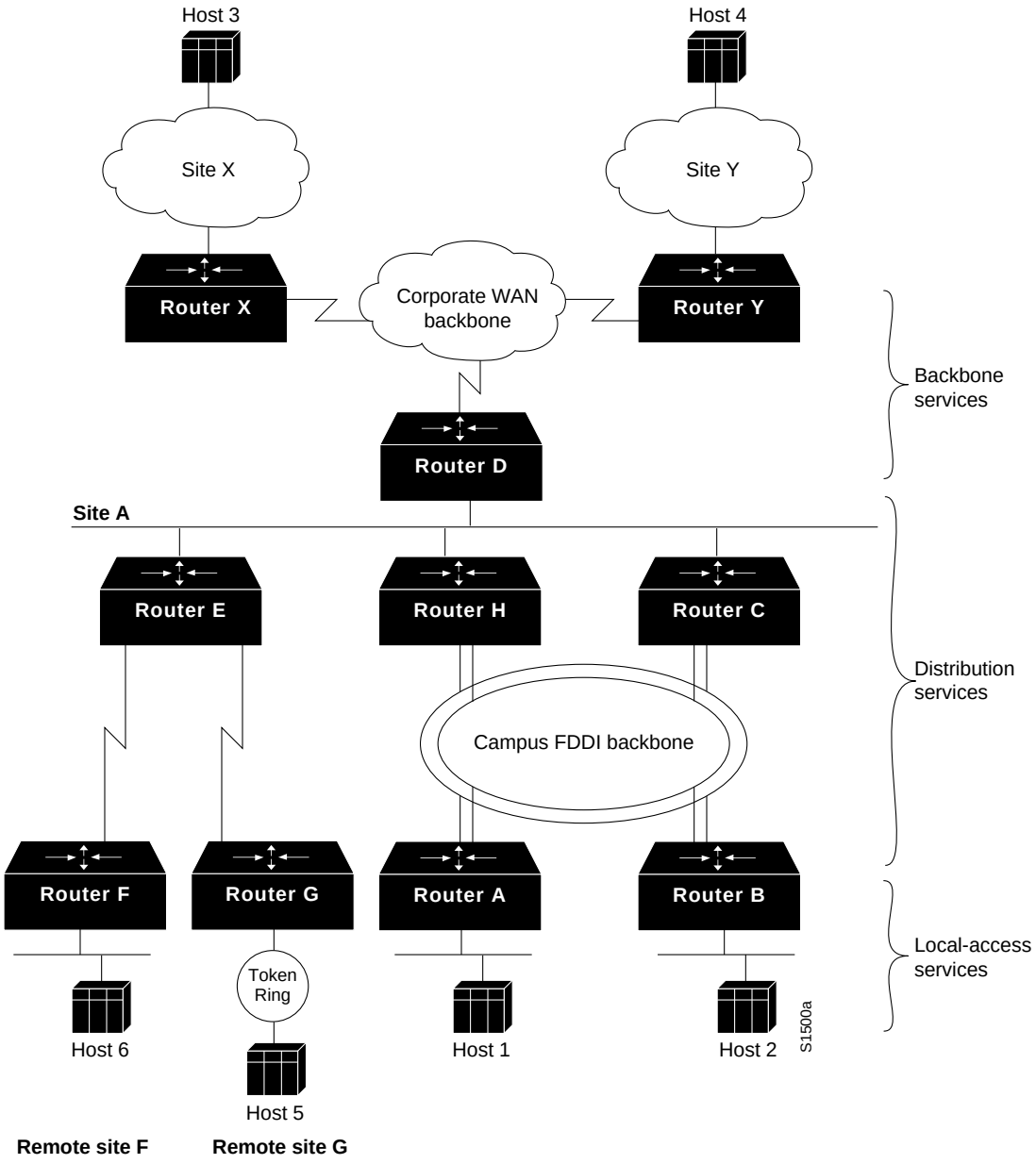
To illustrate the differences between backbone, distribution, and local-access services, consider Figure 1-3. Communication between Host 1 and Router A, or between Host 5 and Router G, is aided by local-access services. Communication between Router D and Router E, or between Router D and Router C, is aided by distribution services. Communication between Router D, Router Y, and Router X is assisted by backbone services.

Assume Host 3 and Host 2 need to communicate. Such a transmission could be routed as follows: Host 3 (through some arbitrary internetworking topology) to Router X to Router D to Router C to Router B to Host 2. The part of the route between Router X and Router D travels through a portion

of the backbone network. Communication between Router D and Router C (and subsequently Router B) is controlled by policy-based rules associated with distribution services. A host and an adjacent router (such as Host 2 and Router B) communicate using local-access services; controlling access to resources on other networks is the primary goal of these local-access routers.

The discussions that follow outline the capabilities and services associated with backbone, distribution, and local access internetworking services.

Figure 1-3 Backbone, Distribution, and Local-Access Router Environment



Evaluating Backbone Services

This section addresses internetworking features that support backbone services. The following topics are discussed:

- Backbone Bandwidth Management
- Path Optimization
- Traffic Prioritization
- Load Balancing
- Alternate Paths
- Switched Access
- Encapsulation (Tunneling)

Backbone Bandwidth Management

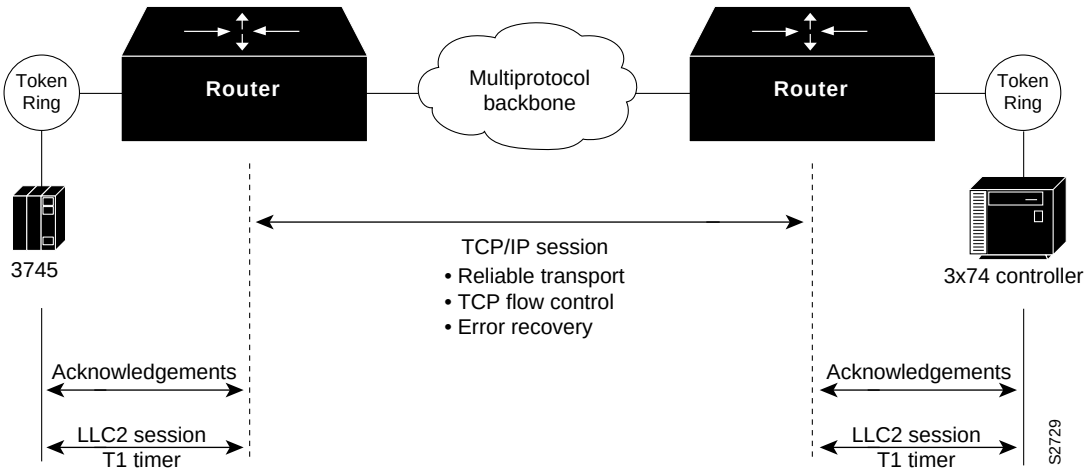
To optimize backbone network operations, routers offer several performance tuning features. Examples include priority queuing, routing protocol metrics, and local session termination.

You can adjust the output queue length on priority queues. If a priority queue overflows, excess packets are discarded and quench messages that halt packet flow are sent, if appropriate, for that protocol.

You can also adjust routing metrics to increase control over the paths that the traffic takes through the internetwork.

Local session termination allows routers to act as proxies for remote systems that represent session endpoints. (A proxy is a device that acts on behalf of another device.) Figure 1-4 illustrates an example of local session termination in an IBM environment.

Figure 1-4 Local Session Termination over Multiprotocol Backbone



In Figure 1-4, the routers locally terminate Logical Link Control type 2 (LLC2) data-link control sessions. Instead of end-to-end sessions, where all session control information is passed over the multiprotocol backbone, the routers take responsibility for acknowledging packets that come from hosts on directly attached LANs. Local acknowledgment saves WAN bandwidth, (and, therefore, WAN utilization costs), solves session timeout problems, and provides faster response to users.

Path Optimization

One of the primary advantages of a router is its ability to help you implement a logical environment in which optimal paths for traffic are automatically selected. Routers rely on routing protocols that are associated with the various network layer protocols to accomplish this automated path optimization.

Depending on the network protocols implemented, routers permit you to implement routing environments that suit your specific requirements. For example, in an IP internetwork, Cisco routers can support all widely implemented routing protocols, including Open Shortest Path First (OSPF), RIP, IGRP, Border Gateway Protocol (BGP), Exterior Gateway Protocol (EGP), and HELLO. Key built-in capabilities that promote path optimization include: rapid and controllable route convergence and tunable routing metrics and timers.

Convergence is the process of agreement, by all routers, on optimal routes. When a network event causes routes to either halt operation or become available, routers distribute routing update messages. Routing update messages permeate networks, stimulating recalculation of optimal routes and eventually causing all routers to agree on these routes. Routing algorithms that converge slowly can cause routing loops or network outages.

Many different metrics are used in routing algorithms. Some sophisticated routing algorithms base route selection on a combination of multiple metrics, resulting in the calculation of a single hybrid metric. IGRP uses one of the most sophisticated distance vector routing algorithms. It combines values for bandwidth, load, and delay to create a composite metric value. Link state routing protocols, such as OSPF and IS-IS, employ a metric that represents the cost associated with a given path.

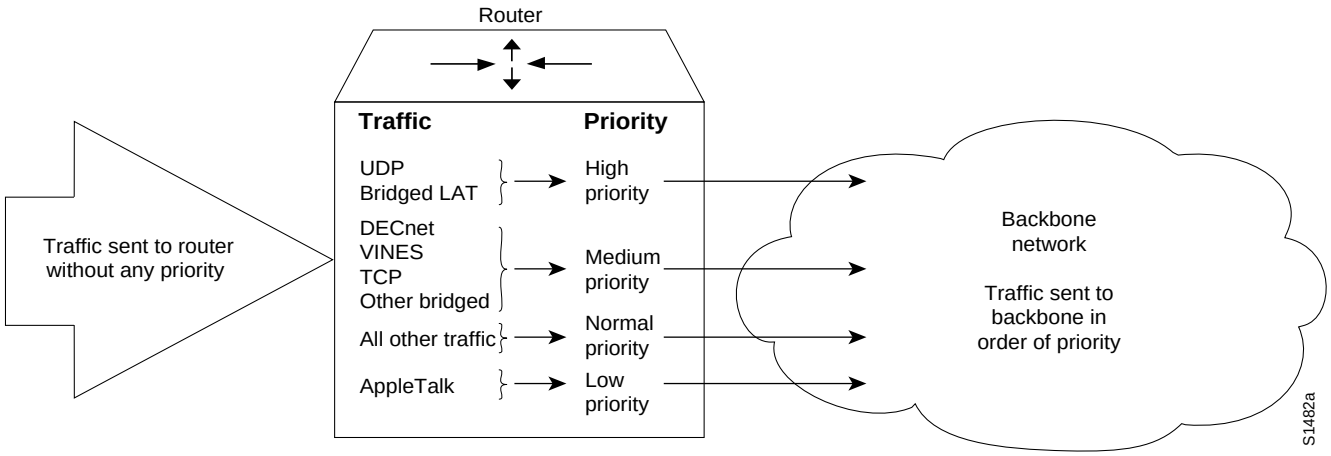
Traffic Prioritization

Although some network protocols can prioritize internal homogeneous traffic, the router prioritizes the heterogeneous traffic flows. Such traffic prioritization enables policy-based routing and ensures that protocols carrying mission-critical data take precedence over less important traffic.

Priority queuing allows the network administrator to prioritize traffic. Traffic can be classified according to various criteria, including protocol and subprotocol type, and then queued on one of four output queues (high, medium, normal, or low priority). For IP traffic, additional fine-tuning is possible.

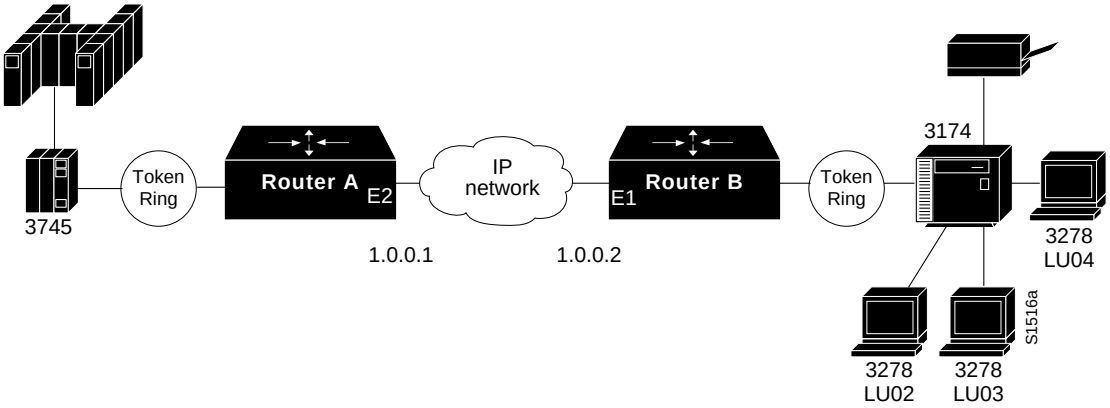
Priority queuing is most useful on low-speed serial links. Figure 1-5 shows how priority queuing can be used to segregate traffic by priority level, speeding the transit of certain packets through the network.

Figure 1-5 Priority Queuing



You can also use intraprotocol traffic prioritization techniques to enhance internetwork performance. IP's type-of-service (TOS) feature and prioritization of IBM logical units (LUs) are intraprotocol prioritization techniques that can be implemented to improve traffic handling over routers. Figure 1-6 illustrates LU prioritization.

Figure 1-6 LU Prioritization Implementation



In Figure 1-6, the IBM mainframe is channel-attached to a 3745 communications controller, which is connected to a 3174 cluster controller via remote source-route bridging (RSRB). Multiple 3270 terminals and printers, each with a unique local LU address, are attached to the 3174. By applying LU address prioritization, you can assign a priority to each LU associated with a terminal or printer; that is, certain users can have terminals that have better response time than others, and printers can have lowest priority. This function increases application availability for those users running extremely important applications.

Finally, most routed protocols (such as AppleTalk, IPX, and DECnet) employ a cost-based routing protocol to assess the relative merit of the different routes to a destination. By tuning associated parameters, you can force particular kinds of traffic to take particular routes, thereby performing a type of manual traffic prioritization.

Load Balancing

The easiest way to add bandwidth in a backbone network is to implement additional links. Routers provide built-in load balancing for multiple links and paths. You can use up to four paths to a destination network. In some cases, the paths need not be of equal cost.

Within IP, routers provide load balancing on both a per-packet and a per-destination basis. For per-destination load balancing, each router uses its route cache to determine the output interface. If IGRP or Enhanced IGRP routing is used, unequal-cost load balancing is possible. The router uses metrics to determine which paths the packets will take; the amount of load balancing can be adjusted by the user.

Load balancing bridged traffic over serial lines is also supported. Serial lines can be assigned to circuit groups. If one of the serial links in the circuit group is in the spanning tree for a network, any of the serial links in the circuit group can be used for load balancing. Data ordering problems are avoided by assigning each destination to a serial link. Reassignment is done dynamically if interfaces go down or come up.

Alternate Paths

Many internetwork backbones carry mission-critical information. Organizations running such backbones are usually interested in protecting the integrity of this information at virtually any cost. Routers must offer sufficient reliability so that they are not the weak link in the internetwork chain. The key is to provide alternate paths that can come on line whenever link failures occur along active networks.

Consider what can go wrong in a simple internetwork. In Figure 1-7, the backbone network consists of the FDDI link between the two corporate buildings as well as serial links A, B, and C, which connect the corporate site to the three remote sites. Secondary networks exist within both the corporate site and the remote sites, but they are not important to this analysis.

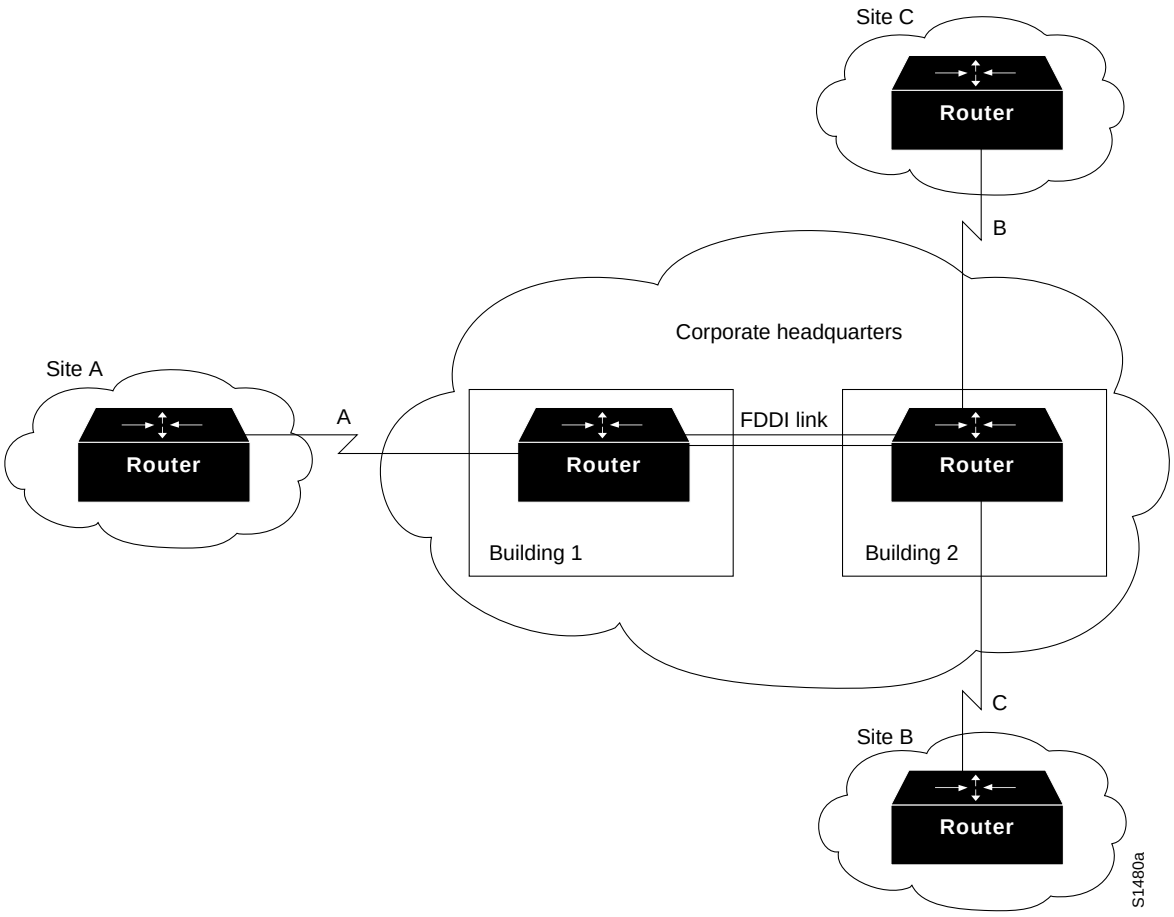
End-to-end reliability is not ensured simply by making the backbone fault tolerant. If communication on a local segment within any building is disrupted for any reason, that information will not reach the backbone. End-to-end reliability is only possible when redundancy is employed throughout the internetwork. Because this is usually cost prohibitive, most companies prefer to employ redundant paths only on those segments that carry mission-critical information.

What does it take to make the backbone reliable? Routers hold the key to reliable internetworking. Depending on the definition of reliability, this can mean duplicating every major system on each router and possibly every component. However, hardware component duplication is not the entire solution because extra circuitry is necessary to link the duplicate components to allow them to communicate. This solution is usually very expensive, but more importantly, it does not completely address the problem. Even assuming all routers in Figure 1-7 are completely reliable systems, link problems between nodes within a backbone can still defeat a redundant hardware solution.

To really address the problem of network reliability, *links* must be redundant. Further, it is not enough to simply duplicate all links. Dual links must terminate at multiple routers unless all backbone routers are completely fault tolerant (no single points of failure). Otherwise, backbone routers that are not fault tolerant become single points of failure. The inevitable conclusion is that a completely redundant router is not the most effective solution to the reliability problem, because it is expensive and still does not address link reliability.

Most network designers do not implement a completely redundant network. Instead, network designers implement partially redundant internetworks. The section, “Choosing Internetworking Reliability Options,” later in this chapter, addresses several hypothetical networks that represent commonly implemented points along the reliability continuum.

Figure 1-7 Simple Internetwork Illustrating Reliability Requirements

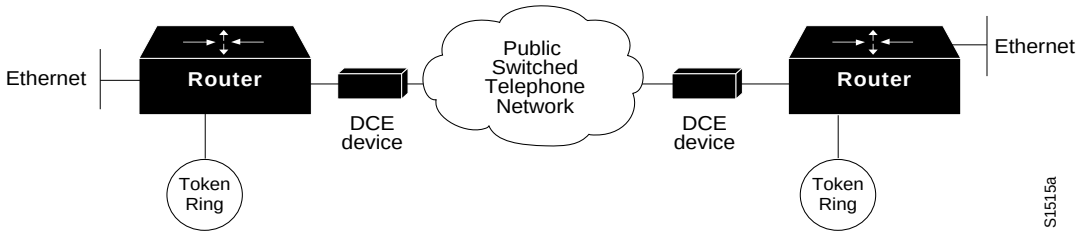


Switched Access

Switched access provides the ability to enable a WAN link on an as-needed basis via automated router controls. One model for a reliable backbone consists of dual, dedicated links and one switched link for idle hot backup. Under normal operational conditions, you can load balance over the dual links, but the switched link is not operational until one of the dedicated links fails.

Traditionally, WAN connections over the Public Switched Telephone Network (PSTN) have used dedicated lines. This can be very expensive when an application requires only low-volume, periodic connections. To reduce the need for dedicated circuits, a feature called dial-on-demand routing (DDR) is available. Using DDR, low-volume, periodic network connections can be made over the PSTN. A router activates the DDR feature when it receives a bridged or routed IP packet destined for a location on the other side of the dial-up line. After the router dials the destination phone number and establishes the connection, packets of any supported protocol can be transmitted. When the transmission is complete, the line is automatically disconnected. By terminating unneeded connections, DDR reduces cost of ownership. Figure 1-8 illustrates a DDR connection.

Figure 1-8 Dial-on-Demand Routing Environment



Encapsulation (Tunneling)

Encapsulation takes packets or frames from one network system and places them inside frames from another network system. This method is sometimes called *tunneling*. Tunneling provides a means for encapsulating packets inside a routable protocol via virtual interfaces. Synchronous Data Link Control (SDLC) transport is also an encapsulation of packets in a routable protocol. In addition, transport provides enhancements to tunneling, such as local data-link layer termination, broadcast avoidance, media conversion, and other scalability optimizations.

Cisco routers support the following encapsulation and tunneling techniques.

- The IBM technology feature set provides these methods:
  - Serial tunneling (STUN) or Synchronous Data Link Control (SDLC) Transport
  - SRB with direct encapsulation
  - SRB with Fast Sequenced Transport (FST) encapsulation
  - SRB with Transmission Control Protocol/Internet Protocol (TCP/IP) encapsulation
- Generic Routing Encapsulation (GRE)

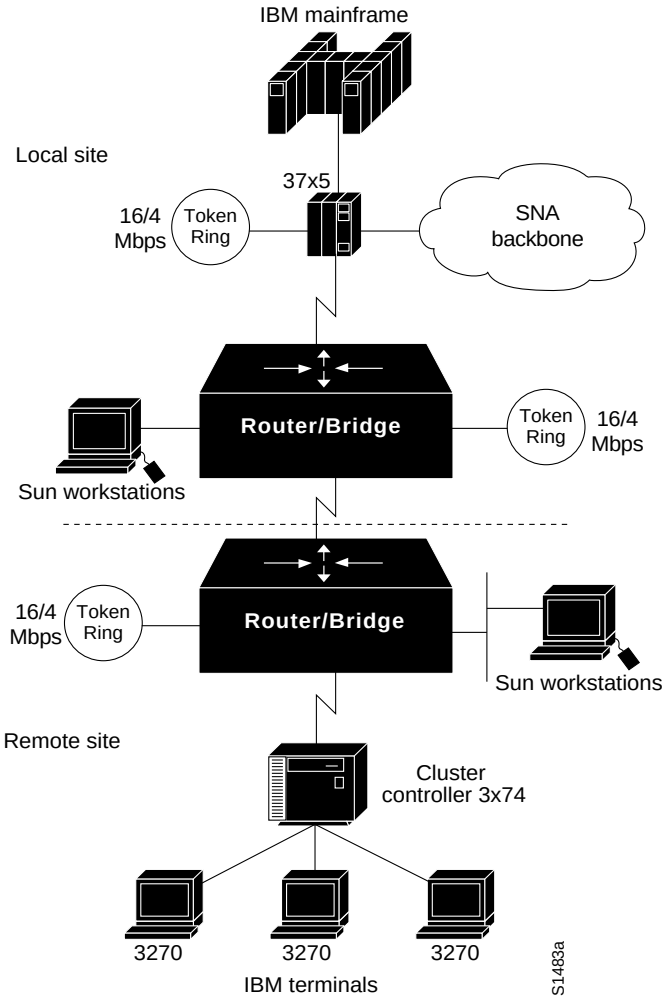
  Cisco supports encapsulating Novell Internetwork Packet Exchange (IPX), Internet Protocol (IP), Connectionless Network Protocol (CLNP), AppleTalk, DECnet Phase IV, Xerox Network Systems (XNS), Banyan Virtual Network System (VINES), and Apollo packets for transport over IP.
- Single-protocol tunneling techniques: Cayman (AppleTalk over IP), AURP (AppleTalk over IP), EON (CLNP over IP), and NOS (IP over IP).

The following discussion focuses on IBM encapsulations and the multiprotocol GRE tunneling feature.

IBM Features

STUN allows two devices that are normally connected by a direct serial link, using protocols compliant with SDLC or High-level Data Link Control (HDLC), to be connected through one or more routers. The routers can be connected via a multiprotocol network of arbitrary topology. STUN allows integration of System Network Architecture (SNA) networks and non-SNA networks using routers and existing network links. Transport across the multiprotocol network that connects the routers can use TCP/IP. This type of transport offers reliability and intelligent routing via any supported IP routing protocol. A STUN configuration is shown in Figure 1-9.

Figure 1-9 STUN Configuration



SDLC Transport is a variation of STUN that allows sessions using SDLC protocols and TCP/IP encapsulation to be locally terminated. SDLC Transport permits participation in SDLC windowing and retransmission activities.

When connecting remote devices that use SRB over a slow-speed serial link, most network designers choose RSRB with direct HDLC encapsulation. In this case, SRB frames are encapsulated in an HDLC-compliant header. This solution adds little overhead, preserving valuable serial link bandwidth. Direct HDLC encapsulation is not restricted to serial links (it can also be used over Ethernet, Token Ring, and FDDI links), but is most useful in situations where additional control overhead on the encapsulating network is not tolerable.

When more overhead can be tolerated, frame sequencing is important, but extremely reliable delivery is not needed, SRB packets can be sent over serial, Token Ring, Ethernet, and FDDI networks using FST encapsulation. FST is similar to TCP in that it provides packet sequencing. However, unlike TCP, FST does not provide packet-delivery acknowledgment.

For extremely reliable delivery in environments where moderate overhead can be tolerated, you can choose to encapsulate SRB frames in TCP/IP packets. This solution is not only reliable, it can also take advantage of routing features that include handling via routing protocols, packet filtering, and multipath routing.

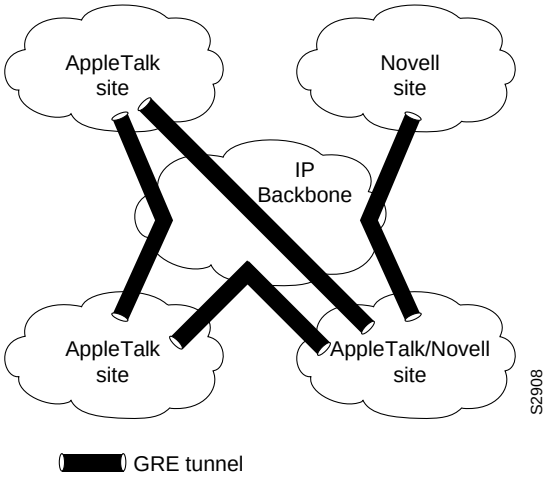
Generic Routing Encapsulation (GRE)

Cisco's Generic Routing Encapsulation (GRE) multiprotocol carrier protocol encapsulates IP, CLNP, IPX, AppleTalk, DECnet Phase IV, XNS, VINES, and Apollo packets inside IP tunnels. With GRE tunneling, a Cisco router at each site encapsulates protocol-specific packets in an IP header, creating a virtual point-to-point link to Cisco routers at other ends of an IP cloud, where the IP header is stripped off. By connecting multiprotocol subnetworks in a single-protocol backbone environment, IP tunneling allows network expansion across a single-protocol backbone environment. GRE tunneling involves three types of protocols:

- Passenger—protocol that is encapsulated (IP, CLNP, IPX, AppleTalk, DECnet Phase IV, XNS, VINES and Apollo)
- Carrier—GRE protocol provides carrier services
- Transport—IP carries the encapsulated protocol

GRE tunneling allows desktop protocols to take advantage of the enhanced route selection capabilities of IP. Many local-area network (LAN) protocols, including AppleTalk and Novell IPX, are optimized for local use. They have limited route selection metrics and hop count limitations. In contrast, IP routing protocols allow more flexible route selection and scale better over large internetworks. Figure 1-10 illustrates GRE tunneling across a single IP backbone between sites. Regardless of how many routers and paths may be associated with the IP cloud, the tunnel is seen as a single hop.

Figure 1-10 Using a Single Protocol Backbone

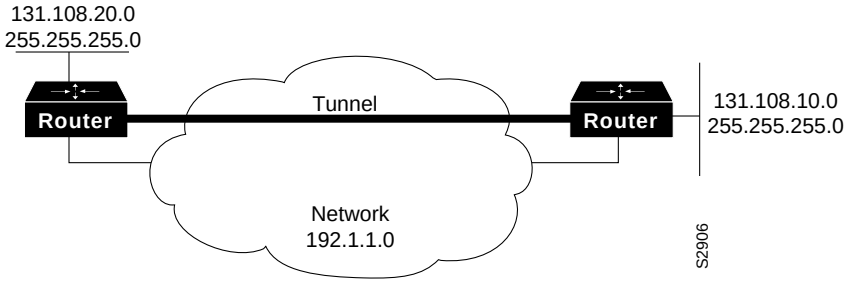


GRE provides key capabilities that other encapsulation protocols lack: sequencing and the ability to carry tunnelled data at high speeds. Some higher-level protocols require that packets are delivered in correct order. The GRE sequencing option provides this capability. GRE also has an optional key feature that allows you to avoid configuration errors by requiring the same key to be entered at each tunnel endpoint before the tunnelled data is processed. IP tunneling also allows network designers

to implement policies, such as which types of traffic can use which routes or assignment of priority or security levels to particular traffic. Capabilities like these are lacking in many native LAN protocols.

IP tunneling provides communication between subnetworks that have invalid or discontiguous network addresses. With tunneling, virtual network addresses are assigned to subnetworks, making discontiguous subnetworks reachable. Figure 1-11 illustrates that with GRE tunneling, it is possible for the two subnetworks of network 131.108.0.0 to talk to each other even though they are separated by another network.

Figure 1-11 Connecting Discontiguous Networks with Tunnels



Because encapsulation requires handling of the packets, it is generally faster to route protocols natively than to use tunnels. Tunneled traffic is switched at approximately half the typical process switching rates. This means approximately 1000 packets per second (pps) aggregate for each router. Tunneling is CPU intensive, and as such, should be turned on cautiously. Routing updates, SAP updates, and other administrative traffic may be sent over each tunnel interface. It is easy to saturate a physical link with routing information if several tunnels are configured over it. Performance depends on the passenger protocol, broadcasts, routing updates, and bandwidth of the physical interfaces. It is also difficult to debug the physical link if problems occur. This problem can be mitigated in several ways. In IPX environments, route filters and SAP filters cut down on the size of the updates that travel over tunnels. In AppleTalk networks, keeping zones small and using route filters can limit excess bandwidth requirements.

Tunneling can disguise the nature of a link, making it look slower, faster, or more or less costly than it may actually be in reality. This can cause unexpected or undesirable route selection. Routing protocols that make decisions based only on hop count will usually prefer a tunnel to a real interface. This may not always be the best routing decision because an IP cloud can comprise several different media with very disparate qualities; for example, traffic may be forwarded across both 100-Mbps Ethernet lines and 9.6-kbps serial lines. When using tunneling, pay attention to the media over which virtual tunnel traffic passes and the metrics used by each protocol.

If a network has sites that use protocol-based packet filters as part of a firewall security scheme, be aware that because tunnels encapsulate unchecked passenger protocols, you must establish filtering on the firewall router so that only authorized tunnels are allowed to pass. If tunnels are accepted from unsecured networks, it is a good idea to establish filtering at the tunnel destination or to place the tunnel destination outside the secure area of your network so that the current firewall scheme will remain secure.

When tunneling IP over IP, you must be careful to avoid inadvertently configuring a recursive routing loop. A routing loop occurs when the passenger protocol and the transport protocol are identical. The routing loop occurs because the best path to the tunnel destination is via the tunnel interface. A routing loop could occur, then, when tunneling IP over IP as follows:

1. The packet is placed in the output queue of the tunnel interface.
2. The tunnel interface includes a GRE header and enqueues the packet to the transport protocol (IP) for the destination address of the tunnel interface.
3. IP looks up the route to the tunnel destination address and learns that the path is the tunnel interface.
4. Once again, the packet is placed in the output queue of the tunnel interface as described in step 1; hence, the routing loop.

When a router detects a recursive routing loop, it shuts down the tunnel interface for 1 to 2 minutes and issues a warning message before it goes into the recursive loop. Another indication that a recursive route loop has been detected is if the tunnel interface is up, and the line protocol is down.

To avoid recursive loops, keep passenger and transport routing information in separate locations by implementing the following procedures:

- Use separate routing protocol identifiers (for example, igrp 1 and igrp 2).
- Use different routing protocols.
- Assign the tunnel interface a very low bandwidth so that routing protocols, such as IGRP, will recognize a very high metric for the tunnel interface and will, therefore, choose the correct next hop (that is, choose the best physical interface instead of the tunnel).
- Keep the two IP address ranges distinct; that is, use a major address for your tunnel network that is different from your actual IP network. Keeping the address ranges distinct also aids in debugging because it is easy to identify an address as the tunnel network instead of the physical network and vice versa.

Evaluating Distribution Services

This section addresses internetworking features that support distribution services. The following topics are discussed:

- Area and Service Filtering
- Policy-Based Distribution
- Gateway Service
- Interprotocol Route Redistribution
- Media Translation

Area and Service Filtering

Traffic filters based on *area* or *service* type are the primary distribution service tools used to provide policy-based access control into backbone services. Both area and service filtering are implemented using *access lists.* An access list is a sequence of statements, each of which either permits or denies certain conditions or addresses. Access lists can be used to permit or deny messages from particular network nodes and messages sent using particular protocols and services.

Area, or network, access filters are used to enforce the selective transmission of traffic based on network address. You can apply these on incoming or outgoing ports. Service filters use access lists applied to protocols (such as IP's UDP), applications such as the Simple Mail Transfer Protocol (SMTP), and specific protocols.

Suppose you have a network connected to the Internet, and you want any host on an Ethernet to be able to form TCP connections to any host on the Internet. However, you do not want Internet hosts to be able to form TCP connections to hosts on the Ethernet except to the SMTP port of a dedicated mail host.

SMTP uses TCP port 25 on one end of the connection and a random port number on the other end. The same two port numbers are used throughout the life of the connection. Mail packets coming in from the Internet will have a destination port of 25. Outbound packets will have the port numbers reversed. The fact that the secure system behind the router always accepts mail connections on port 25 is what makes it possible to separately control incoming and outgoing services. The access list can be configured on either the outbound or inbound interface.

In the following example, the Ethernet network is a Class B network with the address 128.88.0.0, and the mail host's address is 128.88.1.2. The keyword **established** is used only for the TCP protocol to indicate an established connection. A match occurs if the TCP datagram has the ACK or RST bits set, which indicate that the packet belongs to an existing connection.

```
access-list 102 permit tcp 0.0.0.0 255.255.255.255 128.88.0.0 0.0.255.255 established
access-list 102 permit tcp 0.0.0.0 255.255.255.255 128.88.1.2 0.0.0.0 eq 25
interface ethernet 0
ip access-group 102
```

Policy-Based Distribution

Policy-based distribution is based on the premise that different departments within a common organization might have different policies regarding traffic dispersion through the organization-wide internetwork. Policy-based distribution aims to meet the differing requirements without compromising performance and information integrity.

A *policy* within this internetworking context can be defined as a rule or set of rules that govern end-to-end distribution of traffic to (and subsequently through) a backbone network. One department might send traffic representing three different protocols to the backbone, but might wish to expedite one particular protocol's transit through the backbone because it carries mission-critical application information. To minimize already excessive internal traffic, another department might want to exclude all backbone traffic except electronic mail and one key custom application from entering its network segment.

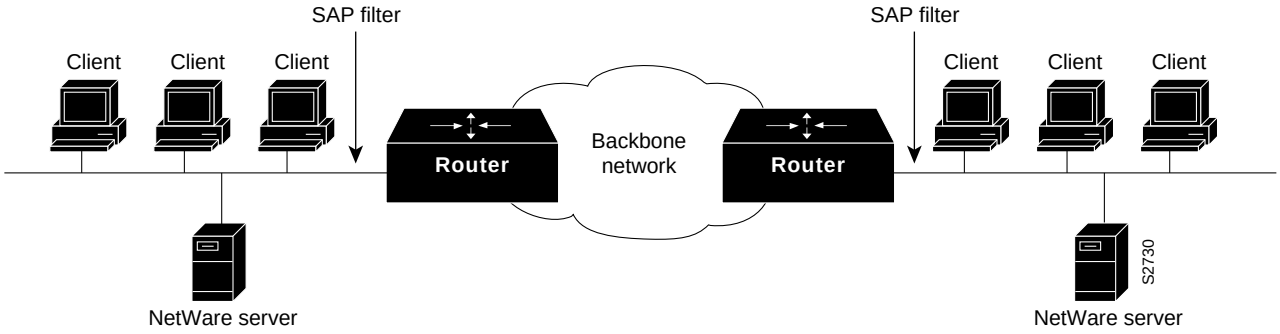
These examples reflect policies specific to a single department. However, policies can reflect overall organizational goals. For example, an organization might want to regulate backbone traffic to a maximum of 10 percent average bandwidth during the work day and 1-minute peaks of 30 percent utilization. Another corporate policy might be to ensure that communication between two remote departments can freely occur, despite differences in technology.

Different policies frequently require different workgroup and department technologies. Therefore, support for policy-based distribution implies support for the wide range of technologies currently used to implement these policies. This in turn allows you to implement solutions that support a wide range of policies, which helps to increase organizational flexibility and application availability.

In addition to support for internetworking technologies, there must be a means both to keep separate and integrate these technologies, as appropriate. The different technologies should be able to coexist or combine intelligently, as the situation warrants.

Consider the situation depicted in Figure 1-12. Assume that a corporate policy limits unnecessary backbone traffic. One way to do this is to restrict the transmission of Service Advertisement Protocol (SAP) messages. SAP messages allow NetWare servers to advertise services to clients. The organization might have another policy stating that all NetWare services should be provided locally. If this is the case, there should be no reason for services to be advertised remotely. SAP filters prevent SAP traffic from leaving a router interface, thereby fulfilling this policy.

Figure 1-12 Policy-Based Distribution: SAP Filtering



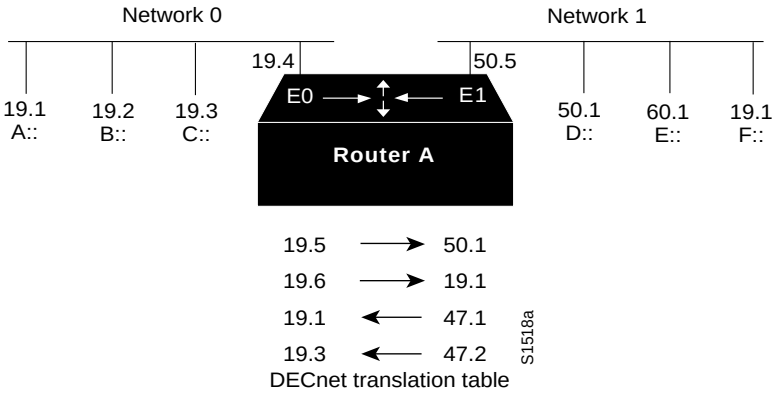
Gateway Service

Protocol gateway capabilities are part of each router's standard software. For example, DECnet is currently in Phase V. DECnet Phase V addresses are different than DECnet Phase IV addresses. For those networks that require both type of hosts to coexist, two-way Phase IV/Phase V translation conforms to Digital-specified guidelines. The routers interoperate with Digital routers, and Digital hosts do not differentiate between the different devices.

The connection of multiple independent DECnet networks can lead to addressing problems. Nothing precludes two independent DECnet administrators from assigning node address 10 to one of the nodes in their respective networks. When the two networks are connected at some later time, conflicts result. DECnet address translation gateways (ATGs) address this problem.

The ATG solution provides router-based translation between addresses in two different DECnet networks connected by a router. Figure 1-13 illustrates an example of this operation.

Figure 1-13 Sample DECnet ATG Implementation



In Network 0, the router is configured at address 19.4 and is a Level 1 router. In Network 1, the router is configured at address 50.5 and is an area router. At this point, no routing information is exchanged between the two networks. The router maintains a separate routing table for each network. By establishing a translation map, packets in Network 0 sent to address 19.5 will be routed to Network 1, and the destination address will be translated to 50.1. Similarly, packets sent to address 19.6 in Network 0 will be routed to Network 1 as 19.1; packets sent to address 47.1 in Network 1 will be routed to Network 0 as 19.1; and packets sent to 47.2 in Network 1 will be sent to Network 0 as 19.3.

AppleTalk is another protocol with multiple revisions, each with somewhat different addressing characteristics. AppleTalk Phase 1 addresses are simple local forms; AppleTalk Phase 2 uses extended (multinetwork) addressing. Normally, information sent from a Phase 2 node cannot be understood by a Phase 1 node if Phase 2 extended addressing is used. Routers support routing between Phase 1 and Phase 2 nodes on the same cable by using transitional routing.

You can accomplish transitional routing by attaching two router ports to the same physical cable. Configure one port to support nonextended AppleTalk and the other to support extended AppleTalk. Both ports must have unique network numbers. Packets are translated and sent out the other port as necessary.

Interprotocol Route Redistribution

The preceding section, "Gateway Service," discussed how *routed* protocol gateways (such as one that translates between AppleTalk Phase 1 and Phase 2) allow two end nodes with different implementations to communicate. Routers can also act as gateways for *routing* protocols. Information derived from one routing protocol such as the IGRP can be passed to and used by another routing protocol such as RIP. This is useful when running multiple routing protocols in the same internetwork.

Routing information can be exchanged between any supported IP routing protocols. These include RIP, IGRP, OSPF, HELLO, EGP, and BGP. Similarly, route redistribution is supported by ISO CLNS for route redistribution between ISO IGRP and IS-IS. Static route information can also be redistributed. Defaults can be assigned so that one routing protocol can use the same metric for all redistributed routes, thereby simplifying the routing redistribution mechanism.

Media Translation

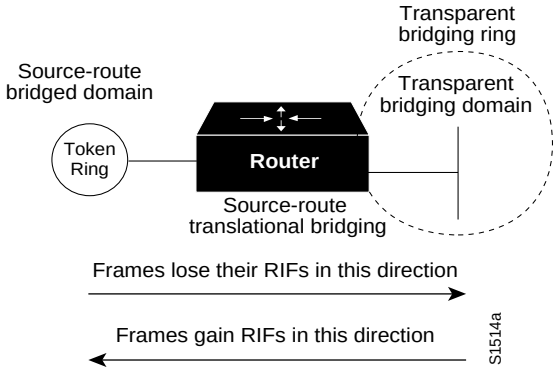
Media translation techniques translate frames from one network system into frames of another. Such translations are rarely 100 percent effective because one system might have attributes with no corollary to the other. For example, Token Ring networks support a built-in priority and reservation system while Ethernet networks do not. Translations between Token Ring and Ethernet networks must somehow account for this discrepancy. It is possible for two vendors to make different decisions about how this discrepancy will be handled, which can prevent multivendor interoperation.

For those situations where communication between end stations on different media is required, routers can translate between Ethernet and Token Ring frames. For direct bridging between Ethernet and Token Ring environments, use either source-route translational bridging or source-route transparent bridging (SRT). Source-route translational bridging translates between Token Ring and Ethernet frame formats; SRT allows routers to use both SRB and the transparent bridging algorithm used in standard Ethernet bridging.

When bridging from the SRB domain to the transparent bridging domain, the SRB fields of the frames are removed. RIFs are cached for use by subsequent return traffic. When bridging in the opposite direction, the router checks the packet to see if it has a multicast or broadcast destination or a unicast destination. If it has a multicast or broadcast destination, the packet is sent as a

spanning-tree explorer. If it has a unicast destination, the router looks up the path to the destination in the RIF cache. If a path is found, it will be used; otherwise, the router will send the packet as a spanning-tree explorer. A simple example of this topology is shown in Figure 1-14.

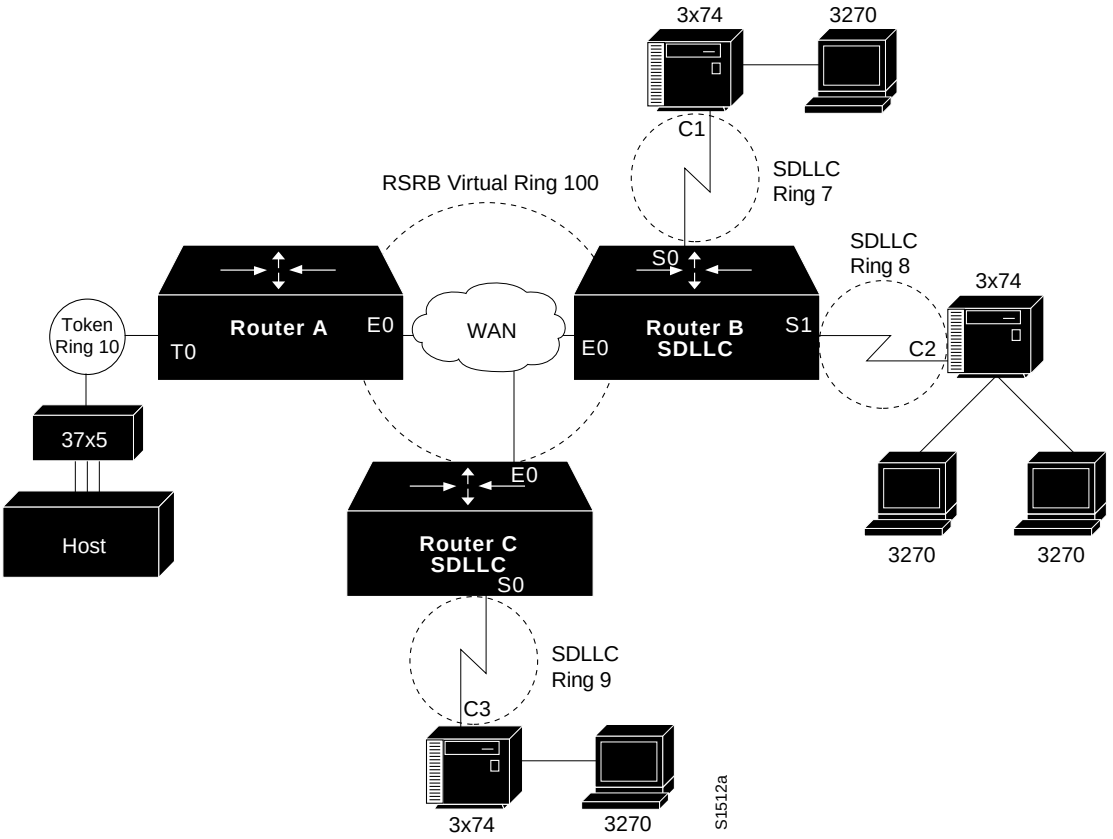
Figure 1-14 Source-Route Translational Bridging Topology



Routers support SRT through implementation of both transparent bridging and SRB algorithms on each SRT interface. If an interface notes the presence of a RIF field, it uses the SRB algorithm; if not, it uses the transparent bridging algorithm.

Translation between serial links running the SDLC protocol and Token Rings running LLC2 is also available. This is referred to as SDLLC frame translation. SDLLC frame translation allows connections between serial lines and Token Rings. This is useful for consolidating traditionally disparate SNA/SDLC networks into a LAN-based, multiprotocol, multimedia backbone network. Using SDLLC, routers terminate SDLC sessions, translate SDLC frames to LLC2 frames, and then forward the LLC2 frames using RSRB over a point-to-point or IP network. Since a router-based IP network can use arbitrary media such as FDDI, Frame Relay, X.25, or leased lines, routers support SDLLC over all such media through IP encapsulation. A complex SDLLC configuration is shown in Figure 1-15.

Figure 1-15 Complex SDLLC Configuration



Evaluating Local-Access Services

The following discussion addresses internetworking features that support local-access services. Local-access service topics outlined here include the following:

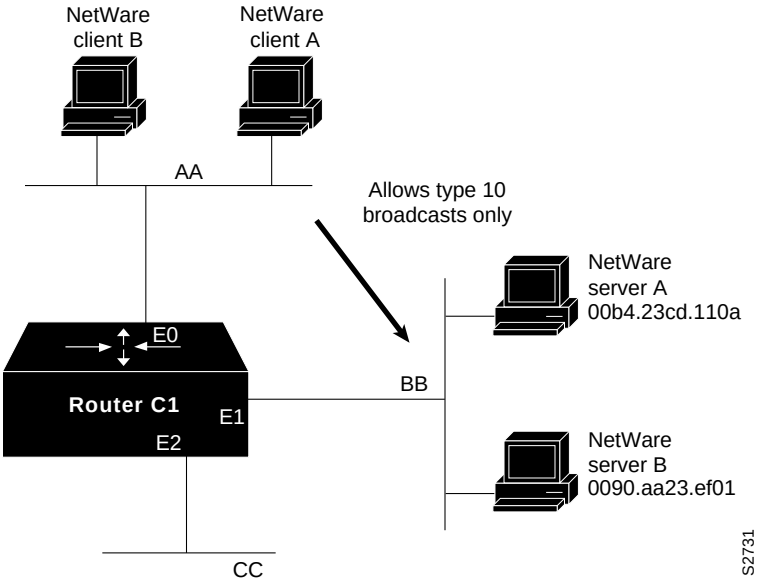
- Value-Added Network Addressing
- Network Segmentation
- Broadcast and Multicast Capabilities
- Naming, Proxy, and Local Cache Capabilities
- Media Access Security
- Router Discovery

Value-Added Network Addressing

Address schemes for LAN-based networks such as NetWare and others do not always adapt perfectly to use over multisegment LANs or WANs. One tool routers implement to ensure operation of such protocols is protocol-specific *helper addressing*. Helper addressing is a mechanism to assist the movement of specific traffic through a network when that traffic might not otherwise transit the network.

The use of *helper addressing* is best illustrated with an example. Consider the use of helper addresses in Novell IPX internetworks. Novell clients send broadcast messages when looking for a server. If the server is not local, broadcast traffic must be sent through routers. Helper addresses and access lists can be used together to allow broadcasts from certain nodes on one network to be directed specifically to certain servers on another network. Multiple helper addresses on each interface are supported, so broadcast packets can be forwarded to multiple hosts. Figure 1-16 illustrates the use of NetWare-based helper addressing.

Figure 1-16 Sample Network Map Illustrating Helper Address Broadcast Control



NetWare clients on Network AA are allowed to broadcast to any server on Network BB. An applicable access list would specify that broadcasts of type 10 will be permitted from all nodes on Network AA. A configuration-specified helper address identifies the addresses on Network BB to which these broadcasts are directed. No other nodes on Network BB receive the broadcasts. No other broadcasts other than type 10 broadcasts are routed.

Any downstream networks beyond Network AA (for example, some Network AA1) are not allowed to broadcast to Network BB through Router C1, unless the routers partitioning Networks AA and AA1 are configured to forward broadcasts with a series of configuration entries. These entries must be applied to the input interfaces and be set to forward broadcasts between directly connected networks. In this way, traffic is passed along in a directed manner from network to network.

Network Segmentation

The splitting of networks into more manageable pieces is an essential role played by local-access routers. In particular, local-access routers implement local policies and limit unnecessary traffic. Examples of capabilities that allow network designers to use local-access routers to segment networks include IP subnets, DECnet area addressing, and AppleTalk zones.

You can use local-access routers to implement local policies by placing the routers in strategic locations and by configuring specific segmenting policies. For example, you can set up a series of LAN segments with different subnet addresses; routers would be configured with suitable interface addresses and subnet masks. In general, traffic on a given segment is limited to local broadcasts,

traffic intended for a specific end station on that segment, or traffic intended for another specific router. By distributing hosts and clients carefully, you can use this simple method of dividing up a network to reduce overall network congestion.

Broadcast and Multicast Capabilities

Many protocols use *broadcast* and *multicast* capabilities. Broadcasts are messages that are sent out to all network destinations. Multicasts are messages sent to a specific subset of network destinations. Routers inherently reduce broadcast proliferation by default. However, routers can be configured to relay broadcast traffic if necessary. Under certain circumstances, passing along broadcast information is desirable and possibly necessary. The key is controlling broadcasts and multicasts using routers.

In the IP world, as with many other technologies, broadcast requests are very common. Unless broadcasts are controlled, network bandwidth can be seriously reduced. Routers offer various broadcast-limiting functions that reduce network traffic and minimize broadcast storms. For example, directed broadcasting allows for broadcasts to a specific network or a series of networks, rather than to the entire internetwork. When flooded broadcasts (broadcasts sent through the entire internetwork) are necessary, Cisco routers support a technique by which these broadcasts are sent over a spanning tree of the network. The spanning tree ensures complete coverage without excessive traffic because only one packet is sent over each network segment.

As discussed previously in the section "Value-Added Network Addressing," broadcast assistance is accommodated with the *helper address* mechanisms. You can allow a router or series of routers to relay broadcasts that would otherwise be blocked by using helper addresses. For example, you can permit retransmission of SAP broadcasts using helper addresses, thereby notifying clients on different network segments of certain NetWare services available from specific remote servers.

The Cisco IP multicast feature allows IP traffic to be propagated from one source to any number of destinations. Rather than sending one packet to each destination, one packet is sent to a multicast group identified by a single IP destination group address. IP multicast provides excellent support for such applications as video and audio conferencing, resource discovery, and stock market traffic distribution.

For full support of IP multicast, IP hosts must run the Internet Group Management Protocol (IGMP). IGMP is used by IP hosts to report their multicast group memberships to an immediately neighboring multicast router. The membership of a multicast group is dynamic. Multicast routers send IGMP query messages on their attached local networks. Host members of a multicast group respond to a query by sending IGMP reports for multicast groups to which they belong. Reports sent by the first host in a multicast group suppress the sending of identical reports from other hosts of the same group.

The multicast router attached to the local network takes responsibility for forwarding multicast datagrams from one multicast group to all other networks that have members in the group. Routers build multicast group distribution trees (routing tables) so that multicast packets have loop-free paths to all multicast group members so that multicast packets are not duplicated. If no reports are received from a multicast group after a set number of IGMP queries, the multicast routers assume the group has no local members and stop forwarding multicasts intended for that group.

Cisco routers support Protocol Independent Multicast (PIM). PIM allows network administrators to add IP multicast routing to their existing IP network regardless of what unicast routing protocol they are using. PIM has two modes: *dense* and *sparse*. In dense mode, receivers are densely populated and the assumption is that networks will probably use the forwarded data. In sparse mode, receivers are widely distributed and the assumption is that the network segment will not use the information.

The Cisco router dynamically discovers Distance Vector Multicast Routing Protocol (DVMRP) neighbors on the interfaces configured for PIM. Once these neighbors are discovered, the Cisco router treats the interface as if there were a PIM neighbor present (with respect to flooding). The Cisco router sends DVMRP Report messages, advertising routes so that sources in the PIM cloud are known to the DVMRP routers. IGMP reports are sent for all groups known in the PIM cloud so that the DVMRP routers know there are group members present.

The Cisco router can also communicate with DVMRP systems using tunnels. You must configure the IP addresses of the end points of the tunnel and indicate that the tunnel operates in DVMRP mode. When DMVRP tunnels are configured, the Cisco router caches DVMRP reports received. This is done so that reverse path forwarding checks are relative to sources reached over the tunnel (without having to send unicast packets over the tunnel). DVMRP tunnels are useful to connect PIM clouds to the MBONE (the Internet Multicast backbone). DVMRP can also be used to connect PIM clouds and MOSPF (Multicast OSPF) clouds together.

Naming, Proxy, and Local Cache Capabilities

Three key router capabilities help reduce network traffic and promote efficient internetworking operation: name service support, proxy services, and local caching of network information.

Network applications and connection services provided over segmented internetworks require a rational way to resolve names to addresses. Various facilities accommodate this requirement. Any router you select must support the name services implemented for different end-system environments. Examples of supported name services include NetBIOS, IP's Domain Name System (DNS) and IEN-116, and AppleTalk Name Binding Protocol (NBP).

A router can also act as a *proxy* for a name server. The router's support of NetBIOS name caching is one example of this kind of capability. NetBIOS name caching allows the router to maintain a cache of NetBIOS names, which avoids the overhead of transmitting all of the broadcasts between client and server NetBIOS PCs (IBM PCs or PS/2s) in an SRB environment.

When NetBIOS name caching is enabled, the router does the following:

- Notices when any host sends a series of duplicate query frames and limits retransmission to one frame per period. The time period is a configuration parameter.
- Keeps a cache of mappings between NetBIOS server and client names and their MAC addresses. As a result, broadcast requests sent by clients to find servers (and by servers in reply to their clients) can be sent directly to their destinations, rather than being broadcast across the entire bridged network.

When NetBIOS name caching is enabled and default parameters are set on the router, the NetBIOS name server, and the NetBIOS name client, approximately 20 broadcast packets per login are kept on the local ring where they are generated.

In most cases, the NetBIOS name cache is best used in situations where large amounts of NetBIOS broadcast traffic might create bottlenecks on a WAN that connects local internetworks to distant locations.

The router can also save bandwidth (or handle nonconforming name resolution protocols) by using a variety of other proxy facilities. By using routers to act on behalf of other devices to perform various functions, you can more easily scale networks. Instead of being forced to add bandwidth when a new workgroup is added to a location, you can use a router to manage address resolution and control message services. Examples of this kind of capability include the proxy explorer feature of SRB and the proxy polling feature of STUN implementations.

Sometimes portions of networks cannot participate in routing activity or do not implement software that conforms to generally implemented address-resolution protocols. Proxy implementations on routers allow network designers to support these networks or hosts without reconfiguring an internetwork. Examples of these kinds of capabilities include proxy ARP address resolution for IP internetworks and NBP proxy in AppleTalk internetworks.

Local caches store previously learned information about the network so that new information requests do not need to be issued each time the same piece of information is desired. A router's ARP cache stores physical address and network address mappings so that it does not need to broadcast ARP requests more than once within a given time period for the same address. Address caches are maintained for many other protocols as well, including DECnet, Novell IPX, and SRB, where RIF information is cached.

Media Access Security

If all corporate information is readily available to all employees, security violations and inappropriate file access can occur. To prevent this, routers must do the following:

- Keep local traffic from inappropriately reaching the backbone
- Keep backbone traffic from exiting the backbone into an inappropriate department or workgroup network

These two functions require packet filtering. Packet filtering capabilities should be tailored to support a variety of corporate policies. Packet filtering methods can reduce traffic levels on a network, thereby allowing a company to continue using its current technology rather than investing in more network hardware. In addition, packet filters can improve security by keeping unauthorized users from accessing information and can minimize network problems caused by excessive congestion.

Routers support many filtering schemes designed to provide control over network traffic that reaches the backbone. Perhaps the most powerful of these filtering mechanisms is the access list. Each of the following possible local-access services can be provided through access lists:

- You have an Ethernet-to-Internet routing network and you want any host on the Ethernet to be able to form TCP connections to any host on the Internet. However, you do not want Internet hosts to be able to form TCP connections into the Ethernet except to the SMTP port of a dedicated mail host.
- You want to advertise only one network through a RIP routing process.
- You want to prevent packets that originated on any Sun workstation from being bridged on a particular Ethernet segment.
- You want to keep a particular protocol based on Novell IPX from establishing a connection between a source network or source port combination and a destination network or destination port combination.

Access lists physically prevent certain packets from traversing a particular router interface, thereby providing a general tool for implementing network security. In addition to this method, several specific security systems already exist to help increase network security. For example, the U.S. government has specified the use of an optional field within the IP packet header to implement a hierarchical packet security system called the Internet Protocol Security Option (IPSO).

IPSO support on routers addresses both the basic and extended security options described in a draft of the IPSO circulated by the Defense Communications Agency. This draft document is an early version of Request for Comments (RFC) 1108. IPSO defines security levels (for example, TOP SECRET, SECRET, and others) on a per-interface basis and accepts or rejects messages based on whether they include adequate authorization.

Some security systems are designed to keep remote users from accessing the network unless they have adequate authorization. For example, the Terminal Access Controller Access Control System (TACACS) is a means of protecting modem access into a network. The Defense Data Network (DDN) developed TACACS to control access to its TAC terminal servers.

The router's TACACS support is patterned after the DDN application. When a user attempts to start an EXEC command interpreter on a password-protected line, TACACS prompts for a password. If the user fails to enter the correct password, access is denied. Router administrators can control various TACACS parameters, such as the number of retries allowed, the timeout interval, and the enabling of TACACS accounting.

The Challenge Handshake Authentication Protocol (CHAP) is another way to keep unauthorized remote users from accessing a network. It is also commonly used to control router-to-router communications. When CHAP is enabled, a remote device (for example, a PC, workstation, router, or communication server) attempting to connect to a local router is "challenged" to provide an appropriate response. If the correct response is not provided, network access is denied.

CHAP is becoming popular because it does not require a secret password to be sent over the network. CHAP is supported on all router serial lines using Point-to-Point Protocol (PPP) encapsulation.

Router Discovery

Hosts must be able to locate routers when they need access to devices external to the local network. When more than one router is attached to a host's local segment, the host must be able to locate the router that represents the optimal path to the destination. This process of finding routers is called *router discovery.*

The following are router discovery protocols:

- End System-to-Intermediate System (ES-IS)—This protocol is defined by the ISO OSI protocol suite. It is dedicated to the exchange of information between intermediate systems (routers) and end systems (hosts). ESs send "ES hello" messages to all ISs on the local subnetwork. In turn, "IS hello" messages are sent from all ISs to all ESs on the local subnetwork. Both types of messages convey the subnetwork and network-layer addresses of the systems that generate them. Using this protocol, end systems and intermediate systems can locate one another.
- ICMP Router Discovery Protocol (IRDP)—Although the issue is currently under study, there is currently no single standardized manner for end stations to locate routers in the IP world. In many cases, stations are simply configured manually with the address of a local router. However, RFC 1256 outlines a router discovery protocol using the Internet Control Message Protocol (ICMP). This protocol is commonly referred to as IRDP.
- Proxy Address Resolution Protocol (ARP)—ARP uses broadcast messages to determine the MAC-layer address that corresponds to a particular internetwork address. ARP is sufficiently generic to allow use of IP with virtually any type of underlying media-access mechanism. A router that has proxy ARP enabled responds to ARP requests for those hosts for which it has a route, which allows hosts to assume that all other hosts are actually on their network.
- RIP—RIP is a routing protocol that is commonly available on IP hosts. Many hosts use RIP to find the address of the routers on a LAN or, when there are multiple routers, to pick the best router to use for a given internetwork address.

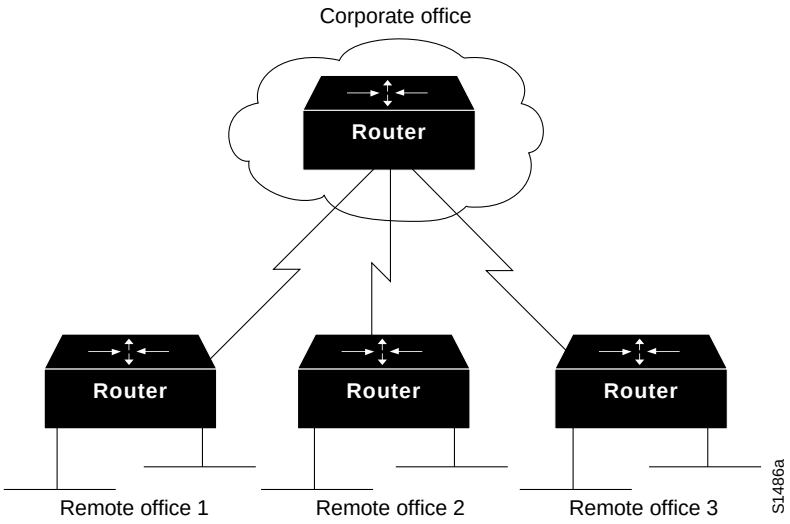
Cisco routers support the router discovery protocols listed above. You can choose the router discovery mechanism that works best in your particular environment.

Choosing Internetworking Reliability Options

The first concern of most network designers is to determine the required level of application availability. In general, this key consideration is balanced against implementation cost. For most organizations, the cost of making a network completely fault tolerant is prohibitive. Determining the appropriate level of fault tolerance to be included in a network, and where redundancy should be used is not trivial.

The nonredundant internetwork design in Figure 1-17 illustrates the considerations involved with increasing levels of internetwork fault tolerance.

Figure 1-17 Typical Nonredundant Internetwork Design



The internetwork shown in Figure 1-17 has two levels of hierarchy: a corporate office and remote offices. Assume the corporate office has 8 Ethernet segments, to which approximately 400 users (an average of 50 per segment) are connected. Each Ethernet segment is connected to a router. In the remote offices, two Ethernet segments are connected to the corporate office through a router. The router in each remote office is connected to the router in the corporate office through a T1 link.

The following sections address various approaches to creating redundant internetworks, provides some context for each approach, and contrasts their relative merits and drawbacks. The following four sections are provided:

- Redundant Links Versus Meshed Topologies
- Redundant Power Systems
- Fault-Tolerant Media Implementations
- Backup Router Hardware

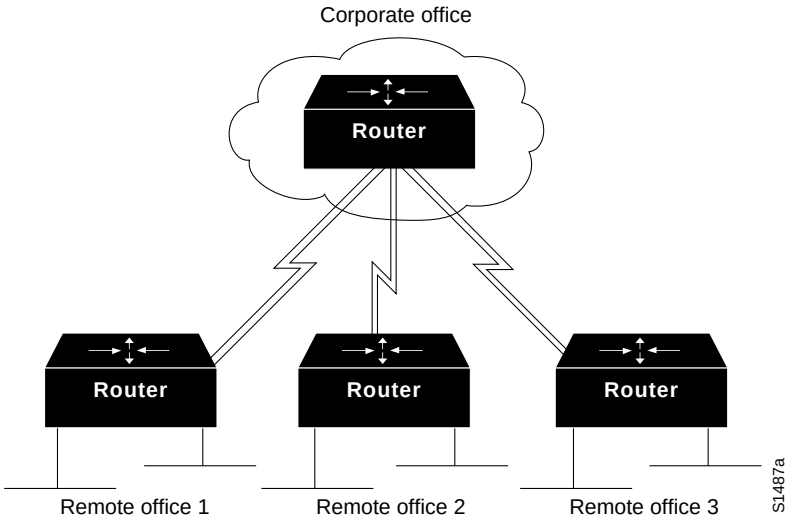
Redundant Links Versus Meshed Topologies

Typically, WAN links are the least reliable components in an internetwork, usually because of problems in the local loop. In addition to being relatively unreliable, these links are often an order of magnitude slower than the LANs they connect. However, because they are capable of connecting

geographically diverse sites, WAN links often make up the backbone network, and are therefore critical to corporate operations. The combination of potentially suspect reliability, lack of speed, and high importance makes the WAN link a good candidate for redundancy.

As a first step in making the example internetwork more fault tolerant, you might add a WAN link between each remote office and the corporate office. This results in the topology shown in Figure 1-18. The new topology has several advantages. First, it provides a backup link that can be used if a primary link connecting any remote office and the corporate office fails. Second, if the routers support load balancing, link bandwidth has now been increased, lowering response times for users and increasing application availability.

Figure 1-18 Internetwork with Dual Links to Remote Offices

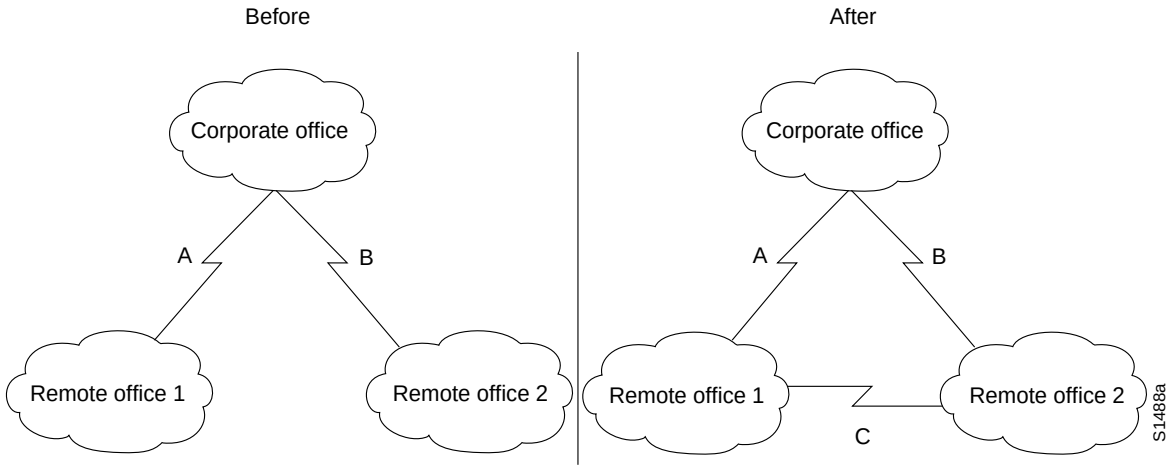


Load balancing in transparent bridging and IGRP environments is another tool for increasing fault tolerance. Routers also support load balancing on either a per-packet or a per-destination basis in all IP environments. Per-packet load balancing is recommended if the WAN links are relatively slow (for example, less than 56 kbps). If WAN links are faster than 56 kbps, enabling fast switching is recommended. When fast switching is enabled, load balancing occurs on a per-destination basis.

Routers can automatically compensate for failed WAN links through routing algorithms of protocols such as IGRP, OSPF, and IS-IS. If one link fails, the routing software recalculates the routing algorithm and begins sending all traffic through another link. This allows applications to proceed in the face of WAN link failure, improving application availability.

The primary disadvantage of duplicating WAN links to each remote office is cost. In the example outlined in Figure 1-18, three new WAN links are required. In large star networks with more remote offices, 10 or 20 new WAN links might be needed, as well as new equipment (including new WAN router interfaces). A lower cost alternative that is becoming increasingly popular links the remote offices using a meshed topology. This topology is illustrated in Figure 1-19.

Figure 1-19 Evolution from a Star to a Meshed Topology



In the "before" portion of Figure 1-19, any failure associated with either Link A or B blocks access to a remote site. The failure might involve the link connection equipment, such as a data service unit (DSU) or a channel service unit (CSU), the router (either the entire router or a single router port), or the link itself. Adding Link C as shown in the "after" portion of the figure, offsets the effect of a failure in any single link. If Link A or B fails, the affected remote site can still access the corporate office through Link C and the other site's link to the corporate office. Note also that if Link C fails, the two remote sites can communicate through their connections to the corporate office.

A meshed topology has three distinct advantages over a redundant star topology:

- A meshed topology is usually slightly less expensive (at least by the cost of one WAN link).
- A meshed topology provides more direct (and, potentially, faster) communication between remote sites, which translates to greater application availability. This can be useful if direct traffic volumes between remote sites are relatively high.
- A meshed topology promotes distributed operation, preventing bottlenecks on the corporate router and further increasing application availability.

A redundant star is a reasonable solution under the following conditions:

- Relatively little traffic must travel between remote offices.
- Traffic moving between corporate and remote offices is delay sensitive and mission critical. The delay and potential reliability problems associated with making an extra hop when a link between a remote office and the corporate office fails might not be tolerable.

Redundant Power Systems

Power faults are common in large-scale networks. Because they can strike across a very local or a very wide scale, power faults are difficult to preempt. Simple power problems include dislodged power cords, tripped circuit breakers, and local power supply failures. More extensive power problems include large-scale outages caused by natural phenomena (such as lightning strikes) or brown-outs. Each organization must assess its needs and the probability of each type of power outage before determining which preventative actions to take.

You can take many precautions to try to ensure that problems such as dislodged power cords do not occur frequently. These fall outside the scope of this publication and will not be discussed here. This publication focuses on issues addressable by internetworking devices.

From the standpoint of internetworking devices, dual power systems can prevent otherwise debilitating failures. Imagine a situation where the so-called "backbone-in-a-box" configuration is being used. This configuration calls for the connection of many networks to a router being used as a "connectivity hub." Benefits include a high-speed backbone (essentially the router's backplane) and cost efficiency (less media). Unfortunately, if the router's power system becomes faulty, each network connected to that router loses its ability to communicate with all other networks connected to that router.

Some backbone-in-a-box routers can address this requirement by providing redundant power systems. In addition, many sites connect one power system to the local power grid and the other to an uninterruptable power supply. If router power fails, the router can continue to provide connectivity to each connected network.

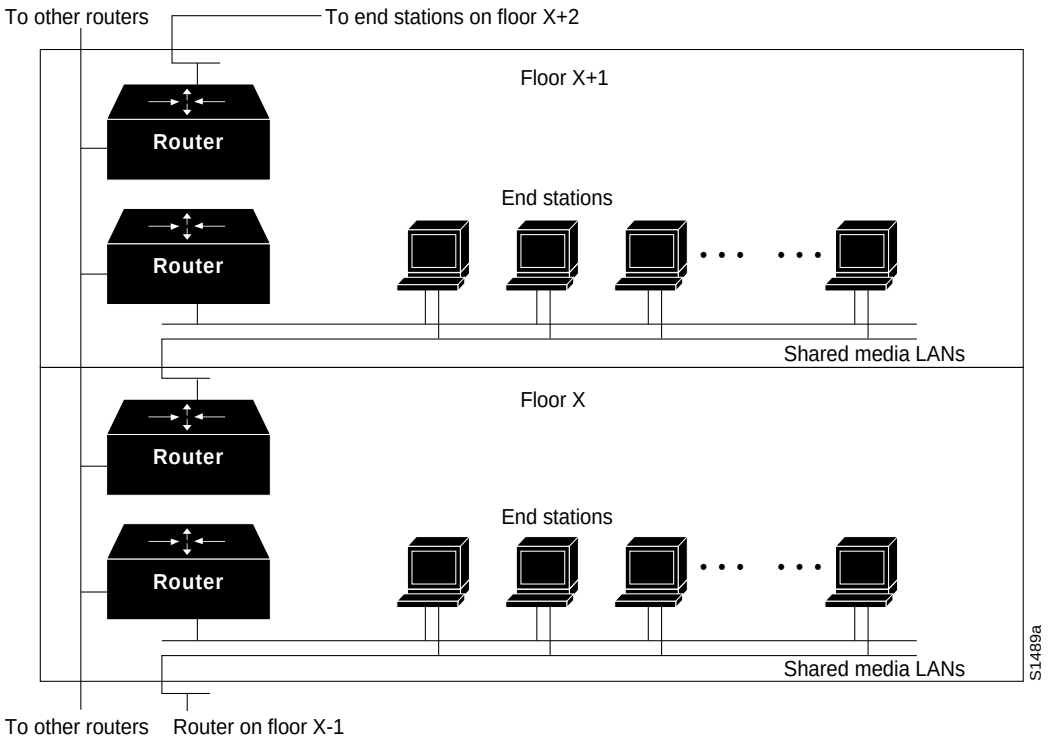
General power outages are usually more common than failures in a router's power system. Consider the effect of a site-wide power failure on redundant star and meshed topologies. If the power fails in the corporate office, the organization might be seriously inconvenienced. Key network applications are likely to be placed at a centralized, corporate location. The organization could easily lose revenue for every minute its network is down. The meshed network configuration is superior in this case, because links between the remote offices would still be able to communicate with each other.

If power fails at a remote site, all connections to that remote site will be terminated, unless otherwise protected. Neither the redundant star nor the meshed topology is superior. In both cases, all other remote offices will still be able to communicate with the corporate office. Generally, power failures in a remote office are more serious when network services are widely distributed.

To protect against local and site-wide power outages, some companies have negotiated an arrangement with local power companies to use multiple power grids within their organization. Failure within one power grid will not affect the network if all critical components have access to multiple power grids. Unfortunately, this arrangement is very expensive and should only be considered by companies with substantial resources, extremely mission-critical operations, and a relatively high likelihood of power failures.

The effect of highly localized power failures can be minimized with prudent network planning. Wherever possible, redundant components should use power supplied by different circuits. Further, these redundant components should not be physically colocated. For example, if redundant routers are employed for all stations on a given floor, these routers can be physically stationed in wiring closets on different floors. This prevents local wiring closet power problems from affecting the ability of all stations on a given floor to communicate. Figure 1-20 shows such a configuration.

Figure 1-20 Redundant Components on Different Floors



For some organizations, the need for fault tolerance is so great that potential power failures are protected against with a duplicate corporate data center. Organizations with these requirements often locate a redundant data center in another city, or in a part of the same city that is some distance from the primary data center. All backend services are duplicated, and transactions coming in from remote offices are sent to both data centers. This configuration requires duplicate WAN links from all remote offices, duplicate network hardware, duplicate servers and server resources, and leasing another building. Because this approach is so costly, it is typically the last step taken by companies desiring the ultimate in fault tolerance.

Partial duplication of the data center is also a possibility. Several key servers and links to those servers can be duplicated. This is a common compromise to the problem presented by power failures.

Fault-Tolerant Media Implementations

Media failure is another possible network fault. Included in this category are all problems associated with the media and its link to each individual end station. Under this definition, media components include network interface controller (NIC) failures, lobe or attachment unit interface (AUI) cable failures, transceiver failures, hub failures, and all failures associated with media components (for example, the cable itself, terminators, and other parts). Many media failures are caused by operator negligence and cannot easily be eliminated.

One way to reduce the havoc caused by failed media is to divide existing media into smaller segments and support each segment with different hardware. This minimizes the effect of a failure on a particular segment. For example, if you have 100 stations attached to a single hub, move some

of them to other hubs. This reduces the effect of a hub failure and of certain subnetwork failures. If you place an internetworking device (such as a router) between segments, you protect against additional problems and cut subnetwork traffic.

As shown in Figure 1-20, redundancy can be employed to help minimize media failures. Each station in this figure is attached to two different media segments. NICs, hub ports, and interface cables are all redundant. This approach doubles the cost of network connectivity for each end station as well as the port usage on all internetworking devices, and is therefore only recommended in situations where complete redundancy is required. It also assumes that end station software, including both the network and the application subsystems, can handle and effectively use the redundant components. The application software or the networking software or both must be able to detect network failures and initiate use of the other network.

Certain media access protocols have some fault-tolerant features built in. Token Ring multistation access units (MAUs) can detect certain media connection failures and bypass the failure internally. FDDI dual rings can wrap traffic onto the backup ring to avoid portions of the network with problems. Partially as a result of their fault tolerant characteristics, use of these protocols is increasing within companies with high reliability needs.

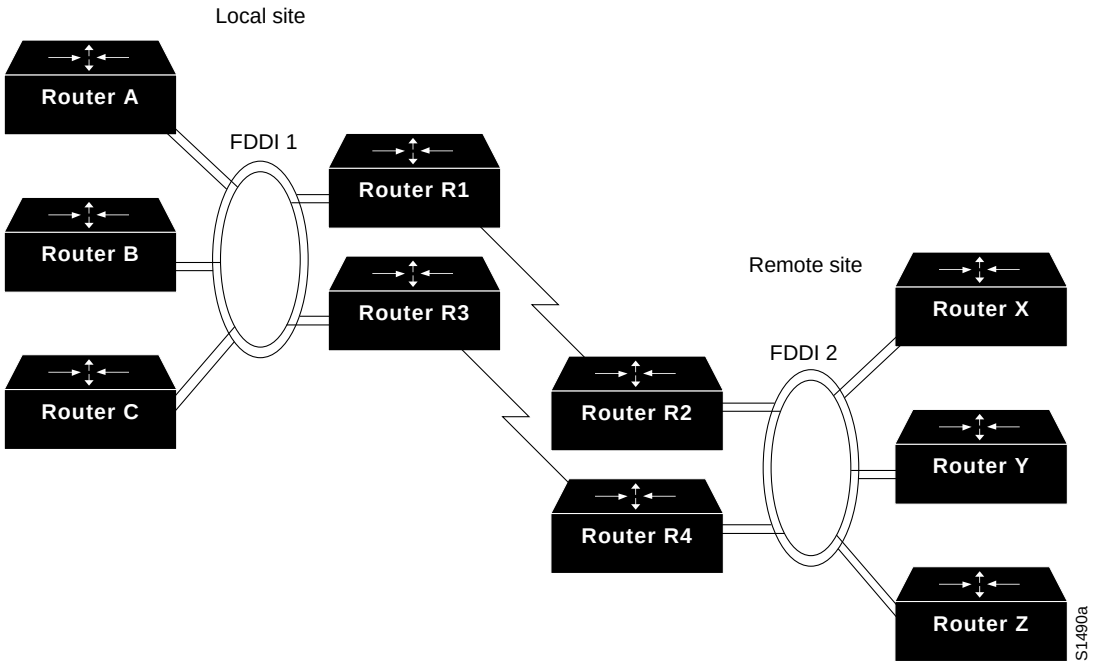
From a router's standpoint, many media failures can be bypassed so long as alternate paths are available. Using various hardware detection techniques, routers can sense certain media-level problems. If routing updates or routing keepalive messages have not been received from devices that would normally be reached through a particular router port, the router will soon declare that route down and will look for alternate routes. Meshed networks provide these alternate paths, allowing the router to compensate for media failures.

Backup Router Hardware

Like all complex devices, routers and other internetworking devices develop hardware problems. Where serious failures occur, the use of dual routers can effectively reduce the adverse effects of a router failure. After a router failure, discovery protocols help end stations choose new local routers with which to communicate. If each network connected to the failed router has an alternate path out of the local area, complete connectivity will still be possible.

When backup routers are used, routing metrics can be set to ensure that the backup routers will not be used unless the primary routers are not functioning. Switchover is automatic and rapid. For example, consider the situation shown in Figure 1-21. In this network, dual routers are used at all sites, with dual WAN links. If Router R1 fails, the routers on FDDI 1 will detect the failure by the absence of messages from Router R1. Using any of several dynamic routing protocols, Router A, Router B, and Router C will designate Router R3 as the new next hop on the way to remote resources accessible via Router R4.

Figure 1-21 Redundant FDDI Router Configuration



Many networks are designed with multiple routers connecting particular LANs in order to provide redundancy. In the past, the effectiveness of this design was limited by the speed at which the hosts on those LANs detected a topology update and changed routers. In particular, IP hosts tend to be configured with a default gateway or configured to use Proxy ARP in order to find a router on their LAN. Convincing an IP host to change its router usually required manual intervention to clear the ARP cache or to change the default gateway.

The Hot Standby Router Protocol (HSRP) is a solution that allows network topology changes to be transparent to the host. HSRP typically allows hosts to reroute in approximately 10 seconds. HSRP is supported on Ethernet, Token Ring, and FDDI. An HSRP group can be defined on each LAN. All members of the group know the standby IP address and the standby MAC address. One member of the group is elected the leader. The lead router services all packets sent to the HSRP group address. The other routers monitor the leader and act as HSRP routers. If the lead router becomes unavailable, the HSRP router elects a new leader who inherits the HSRP MAC address and IP address.

High-end routers (Cisco 7000 and AGS+ families) can support multiple MAC addresses on the same Ethernet or FDDI interface, allowing the routers to simultaneously handle both traffic that is sent to the standby MAC address and the private MAC address.

The commands for enabling HSRP and configuring an HSRP group are **standby ip** and **standby group**.