Windows 2000 Server

Chapter 3 - Unicast IP Routing

Microsoft® Windows® 2000 provides extensive support for unicast Internet Protocol (IP) routing (routing to a unicast destination IP address) with the unicast IP routing protocols and features of the Windows 2000 Router. Your implementation of unicast routing can be simple or complex depending on the size of your IP internetwork, the use of Dynamic Host Configuration Protocol (DHCP) to allocate IP address configuration, connectivity to the Internet, the presence of non-Microsoft or legacy hosts on your internetwork, and other factors.

In This Chapter

Related Information in the Resource Kit

- For more information about unicast routing, see "Unicast Routing Overview" in this book.
- For more information about Windows 2000 and IPX routing, see "IPX Routing" in this book.
- For more information about the Windows 2000 Routing and Remote Access service, see "Routing and Remote Access Service" in this book.

Windows 2000 and IP Routing

An Internet Protocol (IP) router is an IP node that can forward IP packets that are not addressed to the router. Microsoft® Windows NT® version 4.0 and earlier provided a static IP router for simple IP routing. Microsoft® Windows NT® version 3.51 (Service Pack 2 and newer) and Microsoft® Windows NT® Server 4.0 provided a Routing Information Protocol (RIP) for IP service using RIP for IP version 1 (v1). The Routing and Remote Access Service (RRAS) for Windows NT 4.0 (Service Pack 3 and later) provided an integrated IP router with support for RIP for IP (v1 and v2), Open Shortest Path First (OSPF), IP packet filtering, demand-dial routing, and other features.

Windows 2000 Server provides the Routing and Remote Access service with support for RIP for IP, OSPF, IP packet filtering, demand-dial routing, and network address translation. For more information about the components in Windows 2000 that make up the IP router, see "Routing and Remote Access Service" in this book.

Windows 2000 Router Features for IP Routing

The Windows 2000-based computer running the Routing and Remote Access service, known as a Windows 2000 Router, provides a rich set of features to support IP internetworks:

RIP for IP Support for version 1 and version 2 of RIP for IP, the primary routing protocol used in small to medium IP internetworks.

OSPF Support for the industry standard Open Shortest Path First (OSPF) routing protocol used in large and very large IP internetworks.

DHCP Relay Agent Support for an RFC 1542-compliant Dynamic Host Configuration Protocol (DHCP) Relay Agent, also known as a Boot Protocol (BOOTP) Relay Agent, that transfers messages between DHCP clients and DHCP servers located on separate networks.

Network Address Translator (NAT) Support for network address translation to translate private and public addresses to allow the connection of small office or home office (SOHO) networks to the Internet. The network address translator (NAT) component also includes a Dynamic Host Configuration Protocol (DHCP) allocator and a Domain Name System (DNS) proxy to simplify the configuration of SOHO hosts.

IP Packet Filtering Support for separately configured input and output filters for each IP interface based on key fields in the IP, Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and Internet Control Message Protocol (ICMP) headers.

ICMP Router Discovery Support for ICMP Router Advertisements messages to allow the automated discovery of default routers by hosts using ICMP router discovery.

Platform to Support Other IP Routing Protocols Application Programming Interface (API) support that provides a platform to support additional routing protocols such as the Border Gateway Protocol (BGP) for IP. The Windows 2000 Router does not provide BGP, but BGP can be written by third-party independent software vendors. For more information about the API support for third-party routing protocols, see the Windows 2000 Software Development Kit.

Preference Levels

When there are multiple sources of routing information, it becomes necessary to define which route sources are better sources than others. For example, when exchanging routes between RIP and OSPF portions of an intranet, the definition of the metric differs between RIP and OSPF. Rather than trying to reconcile the metrics for two routes to the same destination network ID from different route sources, the route learned from the more preferred route source is used and the route from the less preferred route source is ignored, regardless of the metric.

For example, if a router is configured to use both RIP and OSPF, then both RIP and OSPF-learned routes are added to the Route Table Manager (RTM) IP routing table. If the metric of an OSPF learned route is 5 and the metric of the corresponding RIP learned route is 3 and OSPF is the preferred routing protocol, then the OSPF route is added by RTM to the IP forwarding table.

Preference levels for route sources can be configured from the **Preference Levels** tab on the properties of the **IP Routing\General** container in the **Routing and Remote Access** snap-in. The **Preference Levels** tab allows you to set preference levels for all routes from a specific route source. To set a specific preference level for a static route, use the **netsh routing ip add rtmroute** command.

RIP for IP

RIP for IP is a distance vector routing protocol that facilitates the exchange of IP routing information. Like RIP for IPX, RIP for IP has its origins in the Xerox Network Services (XNS) version of RIP and became a popular routing protocol due to its inclusion in Berkeley UNIX (BSD 4.2 and later) as the *routed* server daemon (a daemon is similar to a Windows 2000 service). There are two versions of RIP. RIP version 1 (v1) is defined in RFC 1058. RIP version 2 (v2) is defined in RFC 1723. In this chapter, information about RIP for IP applies to both versions of RIP. Additional information about the differences between RIP v1 and RIP v2 is also included in this chapter.

RIP and Large Internetworks

While simple and well supported in the industry, RIP for IP suffers from some problems inherent to its original LAN-based design. The combination of the these problems makes RIP a desirable solution only in small to medium-sized IP internetworks.

RIP and Hop Counts

RIP uses a hop count as the metric for the route stored in the IP routing table. The hop count is the number of routers that must be crossed to reach the desired network. RIP has a maximum hop count of 15; therefore, there can only be 15 routers between any two hosts. Networks 16 hops and greater away are considered unreachable. Hop counts can be customized so that slow links are set to multiple hops; however, the accumulated hop count between any two networks must not exceed 15.

The RIP hop count is independent of the Time-to-Live (TTL) field in the IP header. On an internetwork, a network 16 hops away would normally be reachable for an IP packet with an adequate TTL; however, to the RIP router, the network is unreachable and attempts to forward packets to hosts on the network result in ICMP Destination Unreachable-Network Unreachable messages from the RIP router.

RIP and Routing Table Entries

RIP allows for multiple entries in the routing table for a network if there are multiple paths. The IP routing process chooses the route with the lowest metric (lowest hop count) as the best route. However, typical RIP for IP router implementations, including Windows 2000, only store a single lowest metric route for any network. If multiple lowest hop count routes are received by RIP, the first lowest metric route received is stored in the routing table.

If the RIP router is storing a complete list of all the networks and all of the possible ways to reach each network, the routing table can have hundreds or even thousands of entries in a large IP internetwork with multiple paths. Because only 25 routes can be sent in a single RIP packet, large routing tables have to be sent as multiple RIP packets.

RIP Route Advertising

RIP routers advertise the contents of their routing tables every 30 seconds on all attached networks through an IP subnet and MAC-level broadcast. (RIP v2 routers can be configured to multicast RIP announcements.) Large IP internetworks carry the broadcasted RIP overhead of large routing tables. This can be especially problematic on WAN links where significant portions of the WAN link bandwidth are devoted to the passing of RIP traffic. As a result, RIP-based routing does not scale well to large internetworks or WAN implementations.

RIP Convergence

By default, each routing table entry learned through RIP is given a timeout value of three minutes past the last time it was received in a RIP announcement from a neighboring RIP router. When a router goes down due to a hardware or software failure, it can take several minutes for the topology change to be propagated throughout the internetwork. This is known as the slow convergence problem.

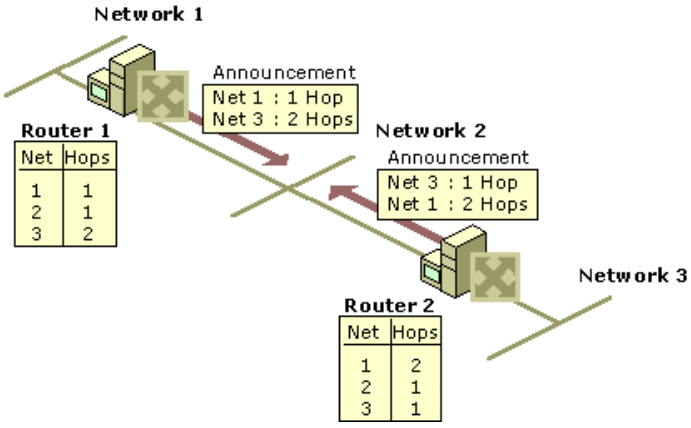
Convergence in RIP Internetworks

RIP for IP, like most distance vector routing protocols, announces its routes in an unsynchronized and unacknowledged manner. This can lead to convergence problems. However, you can enable modifications to the announcement algorithms to reduce convergence time in most situations.

Count-to-Infinity Problem

The classic distance vector convergence problem is known as the count-to-infinity problem and is a direct result of the asynchronous announcement scheme. When RIP for IP routers add routes to their routing table, based on routes advertised by other routers, they keep only the best route in the routing table and they update a lower cost route with a higher cost route only if is being announced by the same source as the current lower cost route. In certain situations, as illustrated in Figures 3.1 through 3.5, this causes the count-to-infinity problem.

Assume that the internetwork in Figure 3.1 has converged. For simplicity, assume that the announcements sent by Router 1 on Network 1 and Router 2 on Network 3 are not included.



If your browser does not support inline frames, click here to view on a separate page.

Figure 3.1 Converged Internetwork

Now assume that the link from Router 2 to Network 3 fails and is sensed by Router 2. As shown in Figure 3.2, Router 2 changes the hop count for the route to Network 3 to indicate that it is unreachable, an infinite distance away. For RIP for IP, infinity is 16.

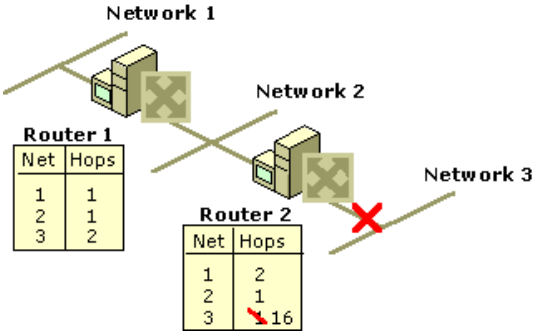


Figure 3.2 Link to Network 3 Fails

However, before Router 2 can advertise the new hop count to Network 3 in a scheduled announcement, it receives an announcement from Router 1. The Router 1 announcement contains a route to Network 3 which is two hops away. Because two hops away is a better route than 16 hops, Router 2 updates its routing table entry for Network 3, changing it from 16 hops to three hops, as shown in Figure 3.3.

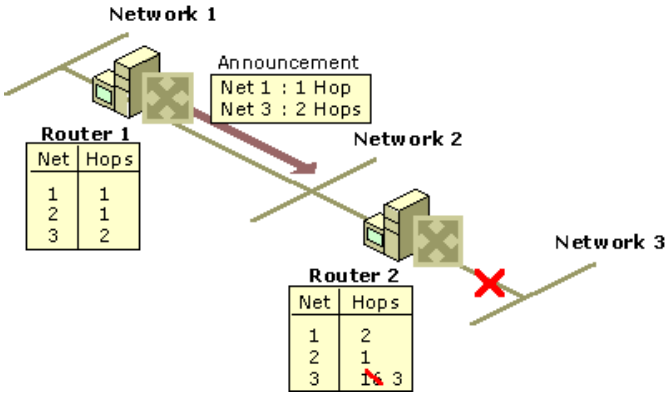


Figure 3.3 Router 2 After Receiving Announcement From Router 1

When Router 2 announces its new routes, Router 1 notes that Network 3 is available three hops away through Router 2. Because the route to Network 3 on Router 1 was originally learned from Router 2, Router 1 updates its route to Network 3 to four hops. (See Figure 3.4.)

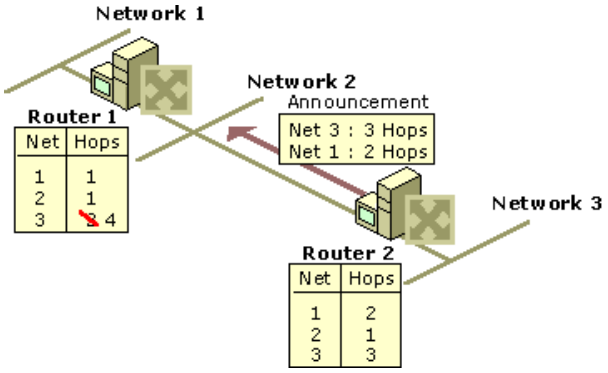


Figure 3.4 Router 1 After Receiving Announcement From Router 2

When Router 1 announces its new routes, Router 2 notes that Network 3 is available four hops away through Router 1. Because the route to Network 3 on Router 2 was originally learned from Router 1, Router 2 updates its route to Network 3 to five hops. (See Figure 3.5.)

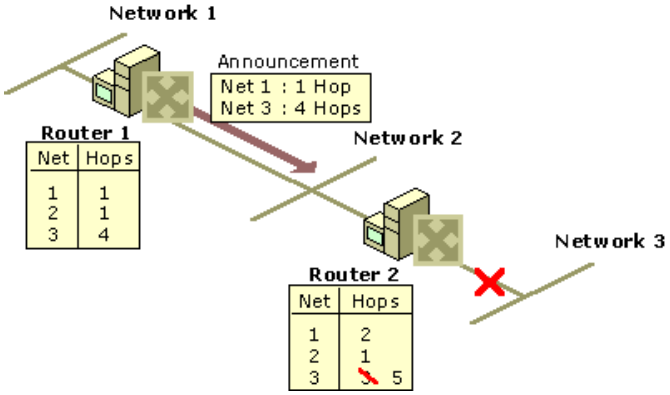


Figure 3.5 Router 2 After Receiving Another Announcement from Router 1

The two routers continue to announce routes to Network 3 with higher and higher hop counts until infinity (16) is reached. Then, Network 3 is considered unreachable and the route to Network 3 is eventually timed out of the routing table. This is known as the count-to-infinity problem.

The count-to-infinity problem is one of the reasons why the maximum hop count of RIP for IP internetworks is set to 15 (16 for unreachable). Higher maximum hop count values would make the convergence time longer when count-to-infinity occurs. Also note that during the count-to-infinity in the previous example, the route from Router 1 to Network 3 is through Router 2. The route from Router 2 to Network 3 is through Router 1. A routing loop exists between Router 1 and Router 2 for Network 3 for the duration of the count-to-infinity problem.

Reducing Convergence Time

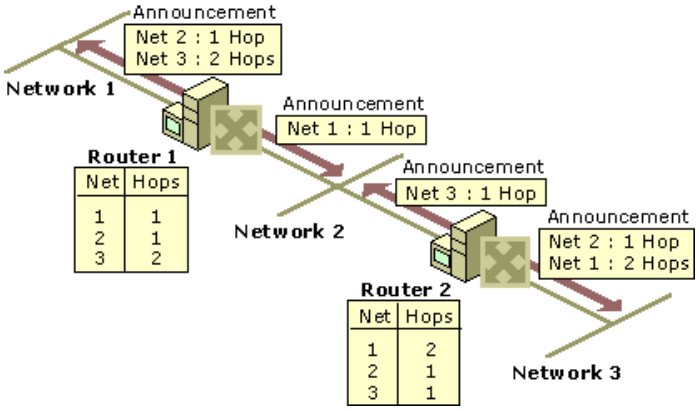
To help reduce the convergence time of RIP for IP internetworks and to avoid count-to-infinity and routing loops in most situations, you can enable the following modifications to the RIP announcement mechanism:

- Split horizon
- Split horizon with poison reverse
- Triggered updates

Split Horizon

Split horizon helps reduce convergence time by not allowing routers to advertise networks in the direction from which those networks were learned. The only information sent in RIP announcements are for those networks that are beyond the neighboring router in the opposite direction. Networks learned from the neighboring router are not included.

Split horizon eliminates count-to-infinity and routing loops during convergence in single-path internetworks and reduces the chances of count-to-infinity in multi-path internetworks. Figure 3.6 illustrates how split horizon keeps the RIP router from advertising routes in the direction from which they were learned.



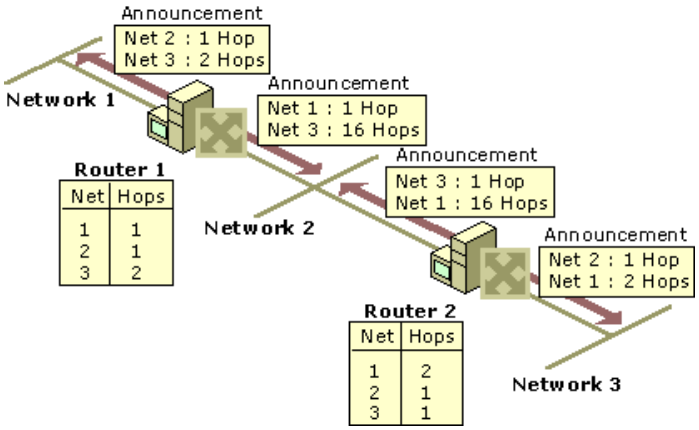
If your browser does not support inline frames, click here to view on a separate page.

Figure 3.6 Split Horizon

Split Horizon with Poison Reverse

Split horizon with poison reverse differs from simple split horizon because it announces all networks. However, those networks learned in a given direction are announced with a hop count of 16, indicating that the network is unreachable. In a single-path internetwork, split horizon with poison reverse has no benefit beyond split horizon. However, in a multipath internetwork, split horizon with poison reverse greatly reduces count-to-infinity and routing loops. Count-to-infinity can still occur in a multipath internetwork because routes to networks can be learned from multiple sources.

In Figure 3.7, split horizon with poison reverse advertises learned routes as unreachable in the direction from which they are learned. Split horizon with poison reverse does have the disadvantage of additional RIP message overhead because all networks are advertised.



If your browser does not support inline frames, click here to view on a separate page.

Figure 3.7 Split Horizon with Poison Reverse

Triggered Updates

Triggered updates allow a RIP router to announce changes in metric values almost immediately rather than waiting for the next periodic announcement. The trigger is a change to a metric in an entry in the routing table. For example, networks that become unavailable can be announced with a hop count of 16 through a triggered update. Note that the update is sent *almost immediately*, where a time interval to wait is typically specified on the router. If triggered updates were sent by all routers immediately, each triggered update could cause a cascade of broadcast traffic across the IP internetwork.

Triggered updates improve the convergence time of RIP internetworks but at the expense of additional broadcast traffic as the triggered updates are propagated.

RIP for IP Operation

The normal operation of a RIP for IP router consists of an initialization process, during which the router learns the routes of the internetwork from neighboring routers; an ongoing periodic advertisement process; and the proper advertisement of unreachable routes when the router is brought down through an administrative action.

Initialization

Upon startup, the RIP for IP router announces its locally attached networks on all of its interfaces. The neighboring RIP routers process the RIP announcement and add the new network or networks to their routing tables as appropriate.

The initializing RIP router also sends a General RIP Request on all locally attached networks. The General RIP Request is a special RIP message requesting all routes. The neighboring RIP routers receive the General RIP Request and send a unicast reply to the requesting router. The replies are used to build the initializing RIP router's routing table.

Ongoing Maintenance

By default, every 30 seconds the RIP router announces it routes on all of its interfaces. The exact nature of the routing announcement depends on whether the RIP router is configured for split horizon or split horizon with poison reverse. The RIP router is also always listening for RIP announcements from neighboring routers in order to add or update the routes in its own routing table.

Administrative Router Shutdown

If a RIP for IP router is downed properly through an administrative action, it sends a triggered update on all locally attached networks. The triggered update announces the networks available through the router with a hop count of 16 (unreachable). This topology change is propagated by neighboring RIP routers throughout the IP internetwork using triggered updates.

As dynamic routers, RIP for IP routers also react to changes in the internetwork topology from downed links or downed routers.

Downed Link

If a link goes down corresponding to one of the router's interfaces and this failure is detected by the interface hardware and indicated to the router, this change is sent out as a triggered update.

Downed Router

If a router goes down due to a power outage or other hardware or software failure, it does not have the ability to inform neighboring routers that the networks available through it have become unavailable. To prevent the lingering existence of unreachable networks in routing tables, each route learned by RIP for IP has a maximum lifetime of 3 minutes (by default). If the entry is not refreshed by the receipt of another announcement in 3 minutes, the entry's hop count is changed to 16 and it is eventually removed from the routing table.

Therefore, if a RIP for IP router goes down, it takes up to 3 minutes for the neighboring routers to mark the routes learned from the downed router as unreachable. Only then do they propagate the topology change throughout the internetwork using triggered updates.

RIP for IP Version 1

RIP version 1 (v1) is defined in RFC 1058 and is widely deployed in small to medium-sized intranets.

RIP v1 Message Format

RIP messages are encapsulated in a User Datagram Protocol (UDP) datagram sent from the router interface IP address and UDP port 520 to the subnet broadcast IP address. The RIP v1 message consists of a 4-byte RIP header and up to 25 RIP routes. The maximum size of the RIP message is 504 bytes. With the 8-byte UDP header, the maximum size of the RIP message is a 512-byte IP payload. Figure 3.8 illustrates the RIP v1 message format.

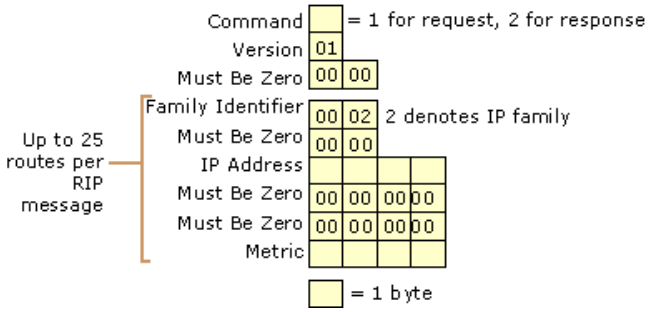


Figure 3.8 RIP Version 1 Message Format

Command A 1-byte field containing either 0x01 or 0x02. 0x01 indicates a RIP request for all (a General RIP Request) or part of the routing tables of neighboring routers. 0x02 indicates a RIP response consisting of all or part of a neighboring router's routing table. A RIP response can be sent in response to a RIP request or as the periodic or triggered update message.

Version A 1-byte field set to the value of 0x01 for RIP v1.

Family Identifier A 2-byte field identifying the protocol family. This is set to the value of 0x00-02 to indicate the IP protocol family.

IP Address A 4-byte field set to the IP network ID which can be a class-based network ID, a subnetted network ID (advertised only within the subnetted network), an IP address (for a host route), or 0.0.0.0 (for the default route). For a General RIP Request, the IP Address is set to 0.0.0.0.

Metric A 4-byte field for the number of hops to the IP network that must be a value from 1 to 16. The metric is set to 16 in a General RIP Request or to indicate that the network is unreachable in a RIP response (announcement).

Problems with RIP v1

RIP v1 was designed in 1988 to suit the dynamic routing needs of LAN technology-based IP internetworks. Shared access LAN technologies like Ethernet and Token Ring support Media Access Control (MAC)-level broadcasting where a single packet can be received and processed by multiple network nodes. However, in modern internetworks, the use of MAC-level broadcasts is undesirable because all nodes must process all broadcasts. RIP v1 was also designed in a time when the Internet was still using network IDs based on the Internet address classes. Today, however, the use of Classless Inter-Domain Routing (CIDR) and variable length subnetting is almost required to conserve IP addresses.

Broadcasted RIP Announcements

All RIP v1 route announcements are addressed to the IP subnet (all host bits are set to 1) and MAC-level broadcast. Non-RIP hosts also receive RIP announcements. For large or very large RIP internetworks, the amount of broadcast traffic on each subnet can become significant.

While producing additional broadcast traffic, the broadcast nature of RIP v1 also permits the use of Silent RIP. A Silent RIP computer processes RIP announcements but does not announce its own routes. Silent RIP could be enabled on non-router hosts to produce a routing table with as much detail as the RIP routers. With more detailed routes in the routing table, a Silent RIP host can make better routing decisions.

Subnet Mask Not Announced with Route

RIP v1 was designed for class-based IP internetworks where the network ID can be determined from the values of the first 3 bits of the IP address in the RIP route. Because the subnet mask is not included with the route, the RIP router must determine the network ID based on a limited set of information. For each route in a RIP v1 message, the RIP v1 router performs the following process:

- If the network ID fits the address classes (Class A, Class B, or Class C), the default class-based subnet mask is assumed.
- If the network ID does not fit the address class, then:
  - If the network ID fits the subnet mask of the interface on which it is received, the subnet mask of the interface on which it was received is assumed.
  - If the network ID does not fit the subnet mask of the interface on which it is received, the network ID is assumed to be a host route with the subnet mask 255.255.255.255.

As a result of the assumptions listed previously, supernetted routes might be interpreted as a single network ID rather than the range of network IDs that they are designed to represent and subnet routes advertised outside of the network ID being subnetted might be interpreted as host routes.

As a mechanism for supporting subnetted environments, RIP v1 routers do not advertise the subnets of a subnetted class-based network ID outside the subnetted region of the IP internetwork. However, because only the class-based network ID is being advertised outside the subnetted environment, subnets of a network ID in a RIP v1 environment must be contiguous. If subnets of an IP network ID are noncontiguous, known as disjointed subnets, the class-based network ID is announced by separate RIP v1 routers in different parts of the internetwork. As a result, IP traffic can be forwarded to the wrong network.

No Protection from Rogue RIP Routers

RIP v1 does not provide any protection from a rogue RIP router starting up on a network and announcing false or inaccurate routes. RIP v1 announcements are processed regardless of their source. A malicious user could use this lack of protection to overwhelm RIP routers with hundreds or thousands of false or inaccurate routes.

RIP for IP Version 2

RIP version 2 (v2) as defined in RFC 1723 seeks to address some of the problems associated with RIP v1. The decision to refine RIP was controversial in the context of newer, smarter routing protocols such as OSPF. However, RIP has the following advantages over OSPF:

- RIP for IP is easy to implement. In its simplest default configuration, RIP for IP is as easy as configuring IP addresses and subnet masks for each router interface and then turning on the router.
- RIP for IP has a large installed base consisting of small and medium-sized IP internetworks that do not wish to bear the design and configuration burden of OSPF.

Features of RIP v2

To help today's IP internetworks minimize broadcast traffic, use variable length subnetting to conserve IP addresses, and secure their routing environment from misconfigured or malicious routers, several key features were added to RIP v2.

Multicasted RIP Announcements

Rather than broadcasting RIP announcements, RIP v2 supports sending RIP announcements to the IP multicast address of 224.0.0.9. Non-RIP nodes are not disturbed by RIP router announcement traffic.

The disadvantage of this new feature is that Silent RIP nodes must also be listening for multicast traffic sent to 224.0.0.9. If you are using Silent RIP, verify that your Silent RIP nodes can listen for multicasted RIP v2 announcements before deploying multicasted RIP v2.

The use of multicasted announcements is optional. The broadcasting of RIP v2 announcements is also supported.

Subnet Masks

RIP v2 announcements send the subnet mask (also known as a network mask) along with the network ID. RIP v2 can be used in subnetted, supernetted, and variable-length subnet mask environments. Subnets of a network ID do not have to be contiguous (they can be disjointed subnets).

Authentication

RIP v2 supports the use of authentication mechanisms to verify the origin of incoming RIP announcements. Simple password authentication was defined in RFC 1723, but newer authentication mechanisms such as Message Digest 5 (MD5) are available.

Note Windows 2000 supports only simple password authentication.

RIP v1 Routers Are Forward Compatible with RIP v2

RIP v1 was designed with forward compatibility in mind. If a RIP v1 router receives a message and the RIP version in the RIP header is not 0x01, it does not discard the RIP announcement but processes only the RIP v1 defined fields.

Also, RIP v2 routers send a RIP v1 response to a RIP v1 request except when configured to send only RIP v2 announcements.

RIP v2 Message Format

To ensure that RIP v1 routers can process RIP v2 announcements, RIP v2 does not modify the structure of the RIP message format. RIP v2 makes use of fields that were defined in RIP v1 as Must be Zero.

The use of the Command, Family Identifier, IP Address, and Metric fields is the same as previously defined for RIP v1. The Version field is set to 0x02 to indicate a RIP v2 message. Figure 3.9 illustrates the RIP v2 message format.

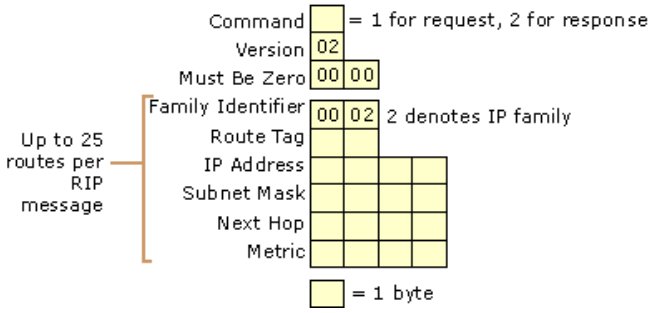


Figure 3.9 RIP Version 2 Message Format

Route Tag

The Route Tag field is used as a method of marking specific routes for administrative purposes. Its original use as defined by RFC 1723 was to distinguish routes that were RIP-based routes (internal to the RIP environment) from non-RIP routes (external to the RIP environment). The Route Tag is configurable on routers that can support multiple routing protocols.

Note Windows 2000 supports the configuration of the Route Tag for RIP v2 interfaces.

Subnet Mask

The 4-byte Subnet Mask field contains the subnet mask (also known as a network mask) of the network ID in the IP Address field.

Next Hop

The 4-byte Next Hop field contains the forwarding IP address (also known as the gateway address) for the network ID in the IP Address field. If the next hop is set to 0.0.0.0, the forwarding IP address (the next hop) for the route is assumed to be the source IP address of the route announcement.

The Next Hop field is used to prevent non-optimal routing situations. For example, if a router announces a host route for a host that resides on the same network as the router interface advertising the route and the Next Hop field is not used, the forwarding IP address for the host route is the IP address of the router's interface, not the IP address of the host. Other routers that receive the announcement on that network forward packets destined for the host's IP address to the announcing router's IP address rather than to the host. This creates a non-optimal routing situation.

Using the Next Hop field, the router announces the host route with the host's IP address in the Next Hop field. Other routers receiving the announcement on that network forward packets destined for the host's IP address to the host's IP address rather than forwarding them to the announcing router.

Because the Next Hop field becomes the Gateway Address field in the IP routing table, the IP address in the Next Hop field should be directly reachable using a router interface.

Authentication in RIP v2

The authentication process for RIP v2 announcements uses the first route entry in the RIP message to store authentication information. The first route entry must be used, leaving a maximum of 24 routes in a RIP v2 authenticated announcement. To indicate authentication, the Family Identifier field is set to 0xFF-FF. The Authentication Type field, normally used as the Route Tag field for a route, indicates the type of authentication being used. Simple password authentication uses the Authentication Type value of 0x00-01.

The 16 bytes after the Authentication Type are used to store the authentication value. For simple password authentication, the 16-byte Authentication Value field stores the left-justified, null-padded, case-sensitive, clear-text password. Figure 3.10 illustrates the RIP v2 authentication message.

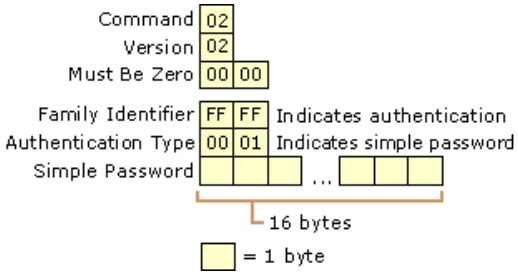


Figure 3.10 RIP v2 Message Format Using Authentication

RIP v1 routers disregard the first route in a RIP v2 authenticated announcement because the Family Identifier for the route is unknown.

Note Simple password authentication for RIP v2 prevents unauthorized or misconfigured RIP routers from being placed on the network. The simple password is not secure, however, because it is sent on the network in clear text. Anyone with a protocol analyzer such as Microsoft Network Monitor can capture the RIP v2 packets and view the authentication password.

Mixed RIP v1 and RIP v2 Environments

RIP v2 routers and RIP v1 routers should be used together with caution. Because RIP v1 routers do not interpret the Subnet Mask field in the route, RIP v2 routers must not announce routes which can be misinterpreted by a RIP v1 router. Variable length subnet masks (VLSM) and disjointed subnets cannot be used in mixed environments.

For an interface using RIP v2 to make announcements such that RIP v1 routers can process the announced routes, the RIP v2 routers must summarize subnet routes when announcing outside a subnetted environment. A specific subnet route announced to a RIP v1 router can be misinterpreted as a host route. Also, the RIP v2 routers cannot announce supernet routes. A RIP v1 router would misinterpret the route as a single network, rather than as a range of networks.

If RIP v2 routers are on the same network as RIP v1 routers, the RIP v2 router interface must be configured to broadcast its announcements. Multicasted RIP v2 announcements are not processed by the RIP v1 routers.

Windows 2000 as a RIP for IP Router

Windows 2000 RIP for IP is RFC 1058 and 1723 compliant and has the following features:

- Split horizon, poison reverse, and triggered updates convergence algorithms.
- Ability to modify the announcement interval (default is 30 seconds).
- Ability to modify the routing table entry timeout value (default is 3 minutes).
- Ability to be a Silent RIP host.
- Peer Filtering: Ability to accept or discard updates of announcements from specific routers identified by IP address.
- Route Filtering: Ability to accept or discard updates of specific network IDs or from specific routers.
- RIP Neighbors: Ability to unicast RIP announcements to specific routers to support nonbroadcast technologies like Frame Relay. A RIP neighbor is a RIP router that receives unicasted RIP announcements.
- Ability to announce or accept default routes or host routes.

Note When a Windows 2000 Router advertises a non-RIP learned route, it advertises it with a hop count of two. Non-RIP learned routes include static routes (even for directly attached networks), OSPF routes, and SNMP routes.

You can view the current RIP neighbors in the **Routing and Remote Access** snap-in by right-clicking the **RIP** routing protocol and clicking **Show Neighbors**.

Troubleshooting RIP for IP

If a RIP environment is properly configured, RIP routers learn all the best routes from neighboring routers after convergence. The exact list of routes added by RIP to the IP routing table depends, among other factors, on whether or not the router interfaces are inside a subnetted region, whether or not RIP v2 is being used, and whether or not host routes or default routes are being advertised.

Problems with RIP can occur in a mixed RIP v1 and v2 environment, with the use of Silent RIP hosts, or when all the appropriate RIP routes are not being received and added to the IP routing table.

Improper routes in a mixed RIP v1 and RIP v2 environment

On networks containing RIP v1 routers, verify that RIP v2 is configured to broadcast its announcements on networks containing RIP v1 routers.

On networks containing RIP v1 routers, verify that the RIP v2 router interfaces are configured to accept both RIP v1 and RIP v2 announcements.

Silent RIP hosts are not receiving routes

If there are Silent RIP hosts on a network that are not receiving routes from the local RIP router, verify the version of RIP supported by the Silent RIP hosts. For example, if the Silent RIP hosts only support listening for broadcasted, RIP v1 announcements, you cannot use RIP v2 multicasting.

If you are using the RIP listener component available on Microsoft® Windows NT® Workstation version 4.0, Service Pack 4 and later, you must configure your RIP routers for RIP v1 or RIP v2 broadcasting.

RIP routers are not receiving expected routes

- Verify that you are not deploying variable length subnetting, disjointed subnets, or supernetting in a RIP v1 or mixed RIP v1 and RIP v2 environment.
- If authentication is enabled, verify that all interfaces on the same network are using the same case-sensitive password.
- If RIP peer filtering is being used, verify that the correct IP addresses for the neighboring peer RIP routers are configured.
- If RIP route filtering is being used, verify that the ranges of network IDs for your internetwork are included or are not being excluded.
- If RIP neighbors are configured, verify that the correct IP addresses are configured for the unicasted RIP announcements.
- Verify that IP packet filtering is not preventing the receiving (through input filters) or sending (through output filters) of RIP announcements on the router interfaces enabled for RIP. RIP traffic uses UDP port 520.
- Verify that TCP/IP filtering on the router interfaces is not preventing the receiving of RIP traffic.
- For dial-up demand-dial interfaces using auto-static updates, configure the demand-dial interfaces to use RIP v2 multicast announcements. When a router calls another router, each router receives an IP address from the other router's IP address pool, which are on different subnets. Because broadcasted RIP announcements are addressed to the subnet broadcast address, each router does not process the other router's broadcasted request for routes. Using multicasting, RIP requests and announcements are processed regardless of the subnet for the router interfaces. For more information about demand-dial interfaces and auto-static updates, see "Demand-Dial Routing" in this book.
- For RIP over demand-dial interfaces, verify that the packet filters on the remote access policy profile of the answering router are not preventing the receipt or sending of RIP traffic. TCP/IP packet filters can be configured on the profile properties of the remote access policies on the answering router (or the Internet Authentication Service (IAS) server if RADIUS is used) that are used to define the traffic that is allowed on the remote access connection.

Host or default routes are not being propagated

- By default, host routes and default routes are not announced using RIP. You can change this behavior from the **Advanced** tab of the properties of a RIP interface in the **Routing and Remote Access** snap-in.

OSPF

Open Shortest Path First (OSPF) is a link-state routing protocol defined in RFC 2328. It is designed to be run as an Interior Gateway Protocol (IGP) to a single Autonomous System (AS). In a link-state routing protocol, each router maintains a database of router advertisements called Link State Advertisements (LSAs). LSAs for routers within the AS consist of a router, its attached networks, and their configured costs. An OSPF cost is a unitless metric that indicates the preference of using a link. There are also LSAs for summarized routes and routes outside of the AS.

The router distributes its LSAs to its neighboring routers. LSAs are gathered into a database called the link state database (LSDB). By synchronizing LSDBs between all neighboring routers, each router has each other router's LSA in its database. Therefore, every router has the same LSDB. From the LSDB, entries for the router's routing table are calculated using the Dijkstra algorithm to determine the least cost path, the path with the lowest accumulated cost, to each network in the internetwork.

OSPF has the following features:

Fast Convergence OSPF can detect and propagate topology changes faster than RIP. Count-to-infinity does not occur with OSPF.

Loop-Free Routes OSPF-calculated routes are always loop-free.

Scalability With OSPF, an AS can be subdivided into contiguous groups of networks called areas. Routes within areas can be summarized to minimize route table entries. Areas can be configured with a default route summarizing all routes outside the AS or outside the area. As a result, OSPF can scale to large and very large internetworks. In contrast, RIP for IP internetworks cannot be subdivided and no route summarization is done beyond the summarizing for all subnets of a network ID.

Subnet Mask Advertised with the Network OSPF was designed to advertise the subnet mask with the network. OSPF supports variable-length subnet masks (VLSM), disjointed subnets, and supernetting.

Support for Authentication Information exchanges between OSPF routes can be authenticated. Windows 2000 OSPF supports simple password authentication.

Support for External Routes Routes outside of the OSPF AS are advertised within the AS so that OSPF routers can calculate the least cost route to external networks.

Note Simple password authentication for OSPF is designed to prevent unauthorized OSPF routers from being placed on the network. The simple password is not secure, however, because it is sent on the network in clear text. Anyone with a protocol analyzer such as Microsoft Network Monitor can capture the OSPF messages and view the authentication password.

OSPF Operation

The main operation of the OSPF protocol occurs in the following consecutive stages and leads to the convergence of the internetwork:

1. Compiling the LSDB.
2. Calculating the Shortest Path First (SPF) Tree.

3. Creating the routing table entries.

Formation of the LSDB Using Link State Advertisements

The LSDB is a database of all OSPF router LSAs, summary LSAs, and external route LSAs. The LSDB is compiled by an ongoing exchange of LSAs between neighboring routers so that each router is synchronized with its neighbor. When the AS has converged, all routers have the appropriate entries in their LSDB.

To create the LSDB, each OSPF router must receive a valid LSA from each other router in the AS. This is performed through a procedure called flooding. Each router initially sends out an LSA which contains its own configuration. As it receives LSAs from other routers, it propagates those LSAs to its neighbor routers.

In this way, an LSA from a given router is flooded across the AS so that each other router contains that router's LSA. While it appears that the flooding of LSAs across the AS causes a large amount of network traffic, OSPF is very efficient in the propagation of LSA information. Figure 3.11 shows a simple OSPF AS, the flooding of LSAs between neighboring routers, and the LSDB.

The exact details of the synchronization of the LSDB between neighboring routers are discussed in the section on creating adjacencies.

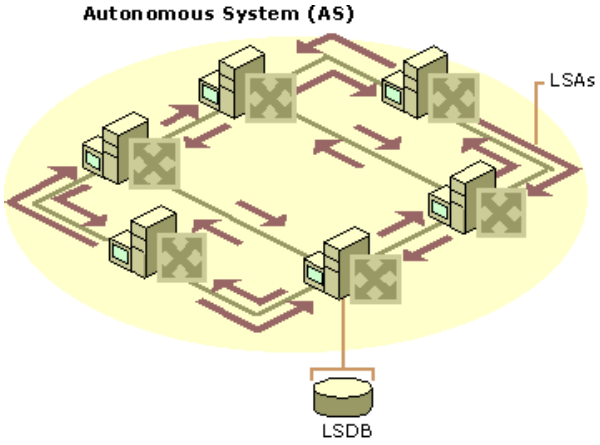


Figure 3.11 OSPF Link State Database (LSDB)

You can view the current OSPF link state database by right-clicking the **OSPF** routing protocol and clicking **Show Link State Database** in the **Routing and Remote Access** snap-in.

Router ID

To keep track of LSAs in the LSDB, each router is assigned a Router ID, a 32-bit dotted decimal number that is unique to the AS. The Router ID identifies the router in the AS, not the IP address of one of the router's interfaces. The Router ID is not used as a destination IP address for sending information to the router. It is a common industry convention to use the largest or smallest IP address assigned to the router as the Router ID. Because IP addresses are unique, this convention ensures that the OSPF Router IDs are also unique.

Calculating the SPF Tree Using Dijkstra's Algorithm

Once the LSDB is compiled, each OSPF router performs a least cost path calculation called the Dijkstra algorithm on the information in the LSDB and creates a tree of shortest paths to each other router and network with themselves as the root. This tree is known as the SPF Tree and contains a single, least cost path to each router and network in the AS. Because the least cost path calculation is performed by each router with itself as the root of the tree, the SPF tree is different for each router in the AS.

The Dijkstra algorithm is from a branch of mathematics called graph theory and is an efficient method of calculating a set of least cost paths relative to a source node.

To Calculate the SPF Tree Using the Dijkstra Algorithm

The result of the Dijkstra algorithm is the set SPF{}, a cost sorted list of least cost paths containing the path (the series of nodes and links) and its accumulated cost from the source node S.

1. Define the set E{} to be the set of nodes (routers) that have been evaluated.
2. Define the set R{} to be the set of nodes (routers) that are remaining (have not been evaluated).
3. Define the set O{} to be a cost-sorted list of ordered paths between nodes. An ordered path can consist of multiple nodes connected together in a multi-hop configuration (they do not have to be neighboring).
4. Define the set SPF{} to be a cost-sorted list of least cost paths containing the path and its accumulated cost.
5. Initialize the set E{} to contain the source node S and the set R{} to contain all other nodes. Initialize the set O{} to be the cost-sorted list of directly connected paths from S. Initialize the set SPF{} to be the empty set.
6. If O{} is empty or the first path in O{} has an infinite metric, mark all the nodes in R as unreachable and terminate the algorithm.
7. From the set O{}, examine P, the shortest ordered path in O{}. Remove P from O{}. Let V be the last node on the ordered path of P.

   If V is already a member of E{}, return to step 6.

   -Or-

   If P is the shortest path to V, move V from R{} to E{}. Add a member to the set SPF{} consisting of P and its accumulated cost from S.
8. Build a new set of paths by concatenating P and each of the links adjacent to V. The cost of these paths is the sum of the cost of P and the metric of the link appended to P. Insert the new links in the set O{} and sort by cost. Return to step 6.

Calculating the Routing Table Entries from the SPF Tree

The OSPF routing table entries are created from the SPF tree, and a single entry for each network in the AS is produced. The metric for the routing table entry is the OSPF-calculated cost, not a hop count.

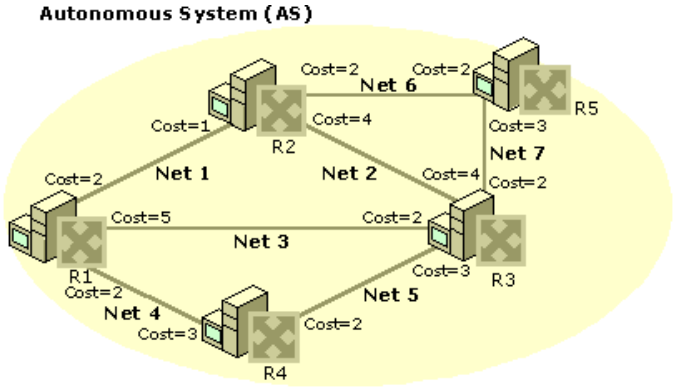
To calculate the IP routing table entries from the SPF Tree, the resulting set SPF{} is analyzed. The result of the analysis is a series of OSPF routes containing the IP destination (the network ID) and its network mask (subnet mask), the forwarding IP address of the appropriate neighboring router, the interface over which the neighboring router is reachable, and the OSPF-calculated cost to the network. The OSPF routes are added to the IP routing table.

Example of OSPF Operation

The following examples illustrate how an OSPF internetwork compiles the LSDB, performs the least cost analysis, and creates routing table entries. This example is deliberately simplified to help you gain an understanding of the basic principles of OSPF convergence.

Compiling the LSDB

Consider the simple AS in Figure 3.12. At each router interface, a unitless cost metric is assigned as a reflection of the preference of using that interface. These cost values can be a reflection of bandwidth, delay, or reliability factors and are assigned by the network administrator.



If your browser does not support inline frames, click here to view on a separate page.

Figure 3.12 AS with Link State Database Information

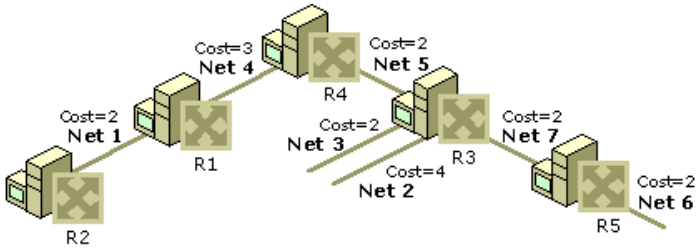
After convergence, when each router has an LSA from each other router in the AS, they each contain the LSDB shown in Table 3.1.

Table 3.1 Link State Database

Router	Attached Networks and Costs
R1	Net 1-Cost 2, Net 3-Cost 5, Net 4-Cost 2
R2	Net 1-Cost 1, Net 2-Cost 4, Net 6-Cost 2
R3	Net 2-Cost 4, Net 3-Cost 2, Net 5-Cost 3, Net 7-Cost 2
R4	Net 4-Cost 3, Net 5-Cost 2
R5	Net 6-Cost 2, Net 7-Cost 3

Calculating the SPF Tree

The next step is to apply Dijkstra's algorithm to our sample OSPF AS. Figure 3.13 illustrates the resulting SPF Tree calculation as performed by router R4. With R4 as the root, the SPF Tree calculation determines the series of connected routers and networks that represent the least cost path to each router and to each network.

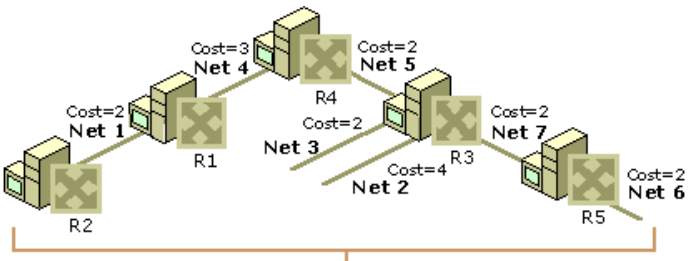


If your browser does not support inline frames, click here to view on a separate page.

Figure 3.13 SPF Tree

Creating Routing Table Entries

The routing table is created from the results of the SPF Tree as shown in Figure 3.14.



Network	Forwarding IP	Port	Metric
4	--	1	3
5	--	2	2
1	R1	1	5
3	R3	2	4
2	R3	2	6
7	R3	2	4
6	R3	2	6

If your browser does not support inline frames, click here to view on a separate page.

Figure 3.14 Routing Table Entries

Note A large OSPF network might require a short but large burst of bandwidth as LSAs are exchanged. Once the LSAs are exchanged a large amount of memory is required to hold the LSDB before convergence. Running the SPF algorithm requires high CPU utilization. Networks with frequently appearing and disappearing links might cause performance issues on any router because of the three-step LSA generation, holding the LSDB, and running the SPF process. The overhead and performance issues of OSPF can be minimized by dividing the OSPF AS into areas. For more information, see "OSPF Areas" later in this chapter.

OSPF Network Types

OSPF message addresses are determined by the type of network to which the OSPF interface is connected. One of the following OSPF network types must be selected when configuring an interface on an OSPF router.

Broadcast A network that can connect more than two routers with a hardware broadcast facility where a single packet sent by a router is received by all routers attached to that network. Ethernet, Token Ring, and FDDI are broadcast networks. OSPF messages sent on broadcast networks use IP multicast addresses.

Point-to-Point A network that can connect only two routers. Leased-line WAN links such as Dataphone Digital Service (DDS) and T-Carrier are point-to-point networks. OSPF messages sent on point-to-point networks use IP multicast addresses.

Non-Broadcast Multiple Access A network that can connect more than two routers but has no hardware broadcast facility. X.25, Frame Relay, and ATM are Non-Broadcast Multiple Access (NBMA) networks. Because multicasted OSPF messages do not reach all the OSPF routers on the network, OSPF must be configured to unicast to the IP addresses of the routers on the NBMA network.

OSPF communication on these network types is discussed in "OSPF Communication on OSPF Networks" later in this chapter.

Note The use of OSPF over nonpermanent, non-persistent, dial-up WAN links such as analog phone lines or ISDN is not recommended.

Synchronizing the LSDB Through Adjacencies

Link-state routing algorithms rely on the synchronization of LSDB information between routers in the AS. Rather than each router verifying synchronization with every other router in the AS, each router is only required to synchronize with its neighboring routers. The relationship between neighboring routers for the purposes of synchronizing the LSDB is called an adjacency. Adjacencies are required to compile the proper entries in the LSDB before the calculation of the SPF Tree and the entries in the routing table. Failure to establish adjacencies is one of the main problems in converging OSPF internetworks. For information about troubleshooting OSPF adjacencies, see "Troubleshooting OSPF" later in this chapter.

Forming an Adjacency

When an OSPF router initializes, it sends out a periodic OSPF Hello packet. The OSPF Hello packet contains configuration information such as the router's Router ID and the list of neighboring routers for which the router has received a Hello packet. Initially, the neighbor list in the OSPF Hello packet does not contain any neighbors.

The initializing OSPF router also listens for neighboring routers' Hello packets. From the incoming Hello packets, the initializing router determines the specific router or routers with which an adjacency is to be established. Adjacencies are formed with the designated router (DR) and backup designated router (BDR) which are identified in the incoming Hello packets. Designated routers and backup designated routers are discussed in more detail later in this chapter.

To begin the adjacency, the routers forming the adjacency describe the contents of their LSDBs through a sequence of Database Description Packets. This is known as the Database Exchange Process during which the two neighboring routers form a master/slave relationship. The contents of each router's LSDB is acknowledged by its neighboring router.

Each router compares its LSAs with the LSAs of its neighbor and notes which LSAs need to be requested from the neighbor to synchronize the LSDB. The missing or more recent LSAs are then requested through Link State Request packets. Link State Update packets are sent in response to the Link State Request packets and their receipt is acknowledged. When all Link State Requests of both routers have been satisfied, the LSDBs of the neighboring routers are fully synchronized and an adjacency is formed.

After the adjacency has formed, each neighboring router sends a periodic Hello packet to inform its neighbor that the router is still active on the network. The lack of Hello packets from a neighbor is used to detect a downed router.

If an event occurs such as a downed link or router or the addition of new network which changes the LSDB of one router, the LSDB of adjacent routers are no longer synchronized. The router whose LSDB has changed sends Link State Update packets to its adjacent neighbor. The receipt of the Link State Update packets is acknowledged. After the exchange, the LSDBs of the adjacent routers are once again synchronized.

Neighbor States

The neighboring routers go through a series of states during the establishment of an adjacency. Table 3.2 lists these states in the adjacency relationship in progressive order.

Table 3.2 Neighbor States for Adjacent Routers

Neighbor	

State	Description
Down	The initial state. No information has been received from the neighbor router.
Attempt	No information has been received despite attempts to contact the neighbor (for NBMA networks only).
Init	A Hello packet has been received from the neighbor, but the router does not appear in the neighbor list of the neighboring router's Hello packet.
2-Way	A Hello packet has been received from the neighbor, and the router does appear in the neighbor list of the neighboring router's Hello packet.
ExStart	Master and slave roles for the Database Exchange Process are being negotiated. This is the first phase of the adjacency relationship.
Exchange	The router is sending Database Description packets to its neighbor.
Loading	Link State Request packets are being sent to the neighbor requesting missing or more recent LSAs.
Full	The neighboring routers' LSDBs are synchronized, and the two routers are fully adjacent.

To view the neighbor state of neighboring routers

- In the **Routing and Remote Access** snap-in, in the **IP Routing** container, right-click **OSPF**, and then click **Show Neighbors**.

  The **OSPF Neighbors** window displays all neighboring routers.

Because of the election of designated routers (DRs) and backup designated routers (BDRs), each OSPF router might not form an adjacency with each other router.

For those neighbor routers where an adjacency is established, **Full** should appear in the State column. For those neighbor routers where an adjacency is not established nor will be established because the router is not a DR or BDR, **2-way** should appear in the State column. Designated routers and backup designated routers are discussed later in this chapter.

Adjacency Configuration Parameters

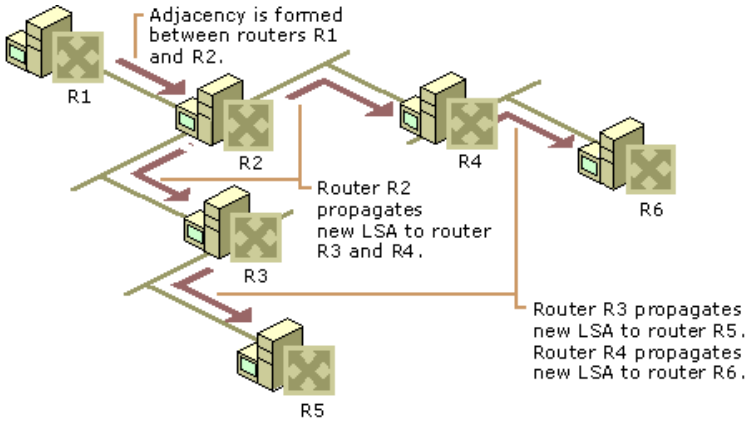
A common OSPF problem is that an adjacency which should form between two neighboring routers does not. The following parameters *must match* between the two routers in order for an adjacency to be established:

- If authentication is being used, the neighboring routers must be using the same authentication type.
- If simple password authentication is enabled, the neighboring routers must be using the same password.
- The Hello Interval (default of 10 seconds), the periodic interval at which Hello packets are sent, must be the same.
- The Dead Interval (default of 40 seconds), the amount of time after which an adjacent router is considered down after ceasing to hear that router's Hello packets, must be the same. The RFC-recommended value is four times the Hello Interval.
- The Area ID, which identifies the area of the AS to which the router is attached, must be the same. The Area ID is configured on each router interface. Areas are discussed in more detail later in this chapter.
- The two neighbor routers must agree as to whether they are in a stub area or not. Stub areas are discussed in more detail later in this chapter.

The Router ID of the two neighboring routers *must not match* in order for an adjacency to be established. Router IDs are designed to be globally unique to the AS. Duplicate Router IDs prevent an adjacency.

Adding a Router to a Converged OSPF Internetwork

When a new OSPF router initializes on an existing and converged OSPF internetwork, the LSA for the new OSPF router must be propagated to all the OSPF routers in the internetwork through flooding. After the new LSA is received, each router must perform Dijkstra's algorithm, recalculate the SPF Tree, and create new routing table entries. After all the routing tables on all the routers are updated, the internetwork has converged.



If your browser does not support inline frames, click here to view on a separate page.

Figure 3.15 New Adjacency Propagation

Figure 3.15 illustrates the convergence process for a new router and new adjacency propagation in a sample OSPF internetwork.

1. Router R1 initializes and begins sending periodic Hello packets across the point-to-point WAN link. Router R2 also sends periodic Hello packets across the link. R1 and R2 decide to form an adjacency.
2. R1 and R2 exchange Database Description Packets. R1's Database Description Packet contains only information about itself. R2's Database Description Packet contains the latest LSAs of all the routers in the internetwork (except R1).
3. R1 sends a Link State Request packet to R2 requesting the LSAs of all the routers on the internetwork. R2 sends the requested LSAs to R1 as Link State Update packets.
4. R2 sends a Link State Request packet to R1 requesting its LSA. R1 sends its LSA to R2 as a Link State Update packet. R1 and R2 now have synchronized LSDBs. Upon receipt of the LSAs, R1 and R2 calculate their respective SPF trees and routing tables.

5. Once synchronized with R1, R2 sends a Link State Update packet to all other OSPF routers to which it is adjacent (routers R3 and R4). The Link State Update packet contains the LSA learned from R1. Upon receipt of the LSA from R2, R3 and R4 calculate their respective SPF trees and routing tables.
6. R3 and R4 flood the information in a separate Link State Update packet to their adjacent routers (routers R5 and R6). Upon receipt of the flooded LSA for R1, R5 and R6 calculate their respective SPF trees and routing tables.

The OSPF internetwork has reconverged after adding R1 and its associated network.

Designated Routers

On a point-to-point link (such as a dedicated WAN link), the adjacency must occur between the two routers on either side of the link. However, on multi-access networks (such as broadcast or NBMA networks) the adjacencies must be controlled. Consider a broadcast network with 6 OSPF routers. Without controlling the adjacency behavior, each router could establish an adjacency with each other router for a total of 15 adjacency relationships. On a broadcast network with *n* routers, a total of $n*(n-1)/2$ adjacencies would be formed. The number of adjacencies scales as $O(n^2)$. In addition, unneeded flooding traffic would occur as each router attempts to synchronize with all of its adjacent routers.

To solve this scaling problem, every multi-access network (broadcast and NBMA) elects a Designated Router (DR). The DR forms adjacencies with all other routers on the network. On a broadcast network with *n* routers, a total of $(n-1)$ adjacencies need to be formed. Because the DR is adjacent with all other routers, it acts as a hub for the distribution of link state information and the LSDB synchronization process.

The DR is elected through the exchange of OSPF Hello packets. Each Hello packet contains the current DR (if elected), the sending router's Router ID, and the sending router's Router Priority. The Router Priority is an interface-specific OSPF configuration parameter that is used to elect the DR. The router with the highest Router Priority is elected the DR. The default Router Priority is 1. A Router Priority of 0 means that the router does not become a DR. If multiple routers have the same highest Router Priority, the router with the highest Router ID is elected the DR.

Caution Router Priorities must be assigned so that at least one router on the multi-access network (broadcast or NBMA) is configured with a Router Priority of 1 or greater. If all routers on a multi-access network are configured with a Router Priority of 0, no router becomes the DR and no adjacencies are established. Without adjacencies, the LSDB cannot be synchronized and no transit traffic (traffic across that network) can be passed.

Note If a DR is already elected for a network, an initializing router on the network does not become the DR if it has a higher Router Priority than the current DR.

DRs on Broadcast Networks

Figure 3.16 illustrates the utility of the DR on a broadcast network. Without a DR, six separate adjacencies would be formed on a broadcast network containing four routers. The redundant adjacencies consume local and network resources maintaining the adjacencies and propagating changes to the LSDB.

With a DR, only three adjacencies are needed to maintain synchronization of the LSDB.

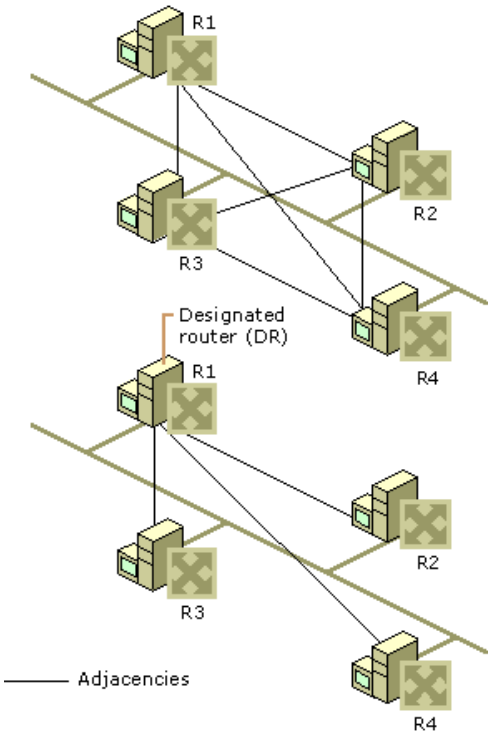


Figure 3.16 Designated Routers on Broadcast Networks

DRs on NBMA Nets

For NBMA networks, such as a Frame Relay network in a hub and spoke configuration, the DR must be the hub router because only the hub router can communicate with all the other routers. To ensure that the hub router is the DR, set its Router Priority to 1 (or greater). To ensure that no spoke router becomes a DR, set spoke router Router Priorities to 0.

Figure 3.17 illustrates a DR in a Frame Relay network.

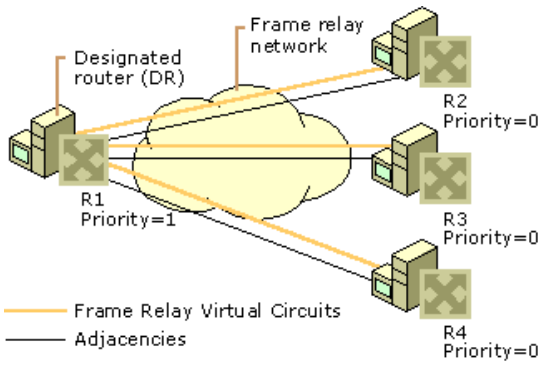


Figure 3.17 Designated Routers on a Frame Relay Network

Backup Designated Router

The DR acts as a central distribution point for topological changes on a multi-access network. If the DR becomes unavailable, all new adjacencies must be formed with a new DR. Until the adjacencies form and the internetwork converges, a temporary loss of connectivity for transit traffic might result.

To prevent the loss in connectivity associated with the loss of a DR, a Backup Designated Router (BDR) is also elected for each multi-access network. Like the DR, the BDR is adjacent to all routers on the network. When the DR fails, the BDR immediately becomes the DR by sending LSAs to all of its adjacent routers announcing its new role. There is a very short period of time where transit traffic could be impaired as the BDR takes over the role of the DR.

Like the DR, the BDR is elected by the exchange of Hello packets. Each Hello packet contains a field for the BDR of the network. If the BDR is not specified, the router with the highest Router Priority that is not already the DR becomes the BDR. If there are multiple routers with the highest Router Priority, the router with the highest Router ID is elected the BDR.

Interface States

Each OSPF interface can be in one of several states after forming adjacencies. Table 3.3 lists the possible interface states.

Table 3.3 Interface States for Adjacent Routers

Interface State	Description
Down	The initial interface state. No Hello packets have been sent or received.
Loopback	The interface to the network is looped back (internally configured so that no packets are sent) through hardware or software.
Waiting	The interface is sending and receiving Hello packets to determine the DR and BDR for the network.
Point-to-Point	The interface is adjacent to its neighbor on a point-to-point network or through a virtual link.
Other	The interface is on a multi-access network and is not the DR or BDR.
Designated Router	The interface is the DR for the attached network.
Backup Designated Router	The interface is the BDR for the attached network.

To view the interface state for an OSPF interface on a Windows 2000 OSPF router, in the **Routing and Remote Access** snap-in, in the **IP Routing** container, click **OSPF**. The contents pane displays the OSPF interfaces. The **State** column indicates the interface's current state. By viewing the interface state for each OSPF interface on a given network, you can determine the DR and the BDR for the network.

OSPF Communication on OSPF Networks

The way in which OSPF routers address OSPF packets varies with the OSPF network type.

Broadcast Networks For broadcast networks, OSPF routers use the following two reserved IP multicast addresses:

- 224.0.0.5 - AllSPFRouters: Used to send OSPF messages to all OSPF routers on the same network. The AllSPFRouters address is used for Hello packets. The DR and BDR use this address to send Link State Update and Link State Acknowledgment packets.
- 224.0.0.6 - AllDRouters: Used to send OSPF messages to all OSPF DRs (the DR and the BDR) on the same network. All OSPF routers except the DR use this address when sending Link State Update and Link State Acknowledgment packets to the DR.

Point-to-Point Networks Point-to-Point networks use the AllSPFRouters address (224.0.0.5) for all OSPF messages.

NBMA Networks NBMA networks have no multicasting capability. Therefore, the destination IP address of any Hello or Link State packets is the unicast IP address of a specific neighbor. The neighbor IP address is a required part of OSPF configuration for NBMA network links.

OSPF Areas

In a very large AS with a large number of networks, each OSPF router must keep the LSA of every other router in its LSDB. Each router in a large OSPF AS has a large LSDB. The SPF calculation of a large LSDB can require a substantial amount of processing. Also, the resulting routing table can be very large, containing a route to each network in the AS.

In an effort to reduce the size of the LSDB and the processing overhead for the SPF tree and routing table calculation, OSPF allows the AS to be divided up into contiguous groups of networks called areas. Areas are identified through a 32-bit Area ID expressed in dotted decimal notation.

An Area ID is an administrative identifier and has no relation to an IP address or IP network ID. Area IDs are not used to reflect routing data. However, if all of the networks within an area correspond to a single subnetted network ID, the area ID can be set to reflect the network ID for administrative convenience. For example, if an area contains all of the subnets of the IP network 10.1.0.0, the area ID can be set to 10.1.0.0.

Reducing the Size of the LSDB

To keep the size of the LSDB for each router to a minimum, LSAs for an area's networks and routers are flooded within the area but not to routers outside the area. Each area becomes its own link state domain with its own LSDB.

If a router is connected to multiple areas, it has multiple LSDBs and SPF Trees. The routing table is a combination of the routing table entries of all the SPF Trees for the router as well as static routes, SNMP configured routes, and routes learned from other routing protocols.

Reducing the Size of the Routing Table

To reduce the number of entries in the routing table of OSPF routers, the networks inside of the area can be advertised outside the area using summary route advertisements. In Figure 3.18, the router on the border of Area 0.0.0.1, known as an area border router (ABR), advertises a summary of all of the networks inside Area 0.0.0.1 in the form of [Destination, Network Mask] pairs to the ABRs of Area 0.0.0.2 and Area 0.0.0.3. Through route summarization, the topology (the networks and their path costs) of an area is hidden from the rest of the AS.

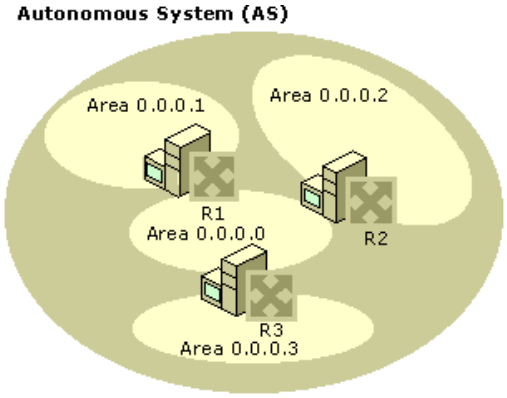


Figure 3.18 OSPF AS and Areas

When the topology of an area is hidden, the rest of the AS is protected from route flapping, events that cause networks to come up or go down. If a network comes up, the event is propagated as a Link State Update and flooded through adjacencies to routers within the area. However, because all the networks within the area are advertised outside the area using summary routes, the Link State Update is not flooded outside the area.

You can view the current OSPF areas by right-clicking the **OSPF** routing protocol and clicking **Show Areas** in the **Routing and Remote Access** snap-in.

Backbone Area

An OSPF internetwork, whether or not it is subdivided into areas, always has at least one area called the backbone. The backbone has the reserved area ID of 0.0.0.0. The OSPF backbone area is also known as area 0.

The backbone acts as a hub for inter-area transit traffic and the distribution of routing information between areas. Inter-area traffic is routed to the backbone, then routed to the destination area, and finally routed to the destination host within the destination area (for more information, see "Inter-Area Routing" later in this chapter). Routers on the backbone also advertise the summarized routes within their areas to the other routers on the backbone. These summary advertisements are flooded into area routers. Therefore, each router in an area has a routing table that reflects the routes available within its area and the routes corresponding to the summary advertisements of the ABRs of the other areas in the AS.

For example, in Figure 3.18, router R1 advertises all of the routes (the list of address ranges) in Area 0.0.0.1 to all backbone routers (routers R2 and R3) using a summary advertisement. R1 receives summary advertisements from R2 and R3. R1 is configured with summary advertisement information for Area 0.0.0.0. Through flooding, R1 propagates that summary routing information to all of the routers within Area 0.0.0.1. For each router within Area 0.0.0.1, the summary routing information from Areas 0.0.0.0, 0.0.0.2, and 0.0.0.3 is incorporated into the calculation of the routing table.

OSPF Router Types

When an OSPF AS is subdivided into areas, the routers are classified by one or more of the categories defined in Table 3.4.

Table 3.4 OSPF Router Types

Router Type	Description
Internal Router	A router with all interfaces connected to the same area. Internal routers each have a single LSDB.
Area Border Router (ABR)	A router with interfaces connected to different areas. ABRs have multiple LSDBs, one for each attached area.
Backbone Router	A router with an interface on the backbone area. This includes all ABRs and internal routers of the backbone area.
AS Boundary Router (ASBR)	A router that exchanges routes with sources outside of the OSPF AS. ASBRs advertise external routes throughout the OSPF AS.

Inter-Area Routing

Routing within an area is performed by OSPF routers using the least cost path to the destination network. Because routes within an area are not summarized, each router has a route to each network within its area or areas.

Routing between areas takes the following course:

1. Routers within the source area forward the packet along the least cost path to the nearest ABR.
2. Backbone routers forward the packet along the least cost path to the nearest ABR connected to the area containing the IP address of the destination host.
3. Routers within the area containing the IP address of the destination host forward the packet along the least cost path to the destination host.

In Figure 3.19, the packet is forwarded across the routers of Area 0.0.0.1 to router R1, a backbone area router. Then, the packet is forwarded across the routers of the backbone area (Area 0.0.0.0) to Router R2. And finally, the packet is forwarded across the routers of Area 0.0.0.2 to the destination host.

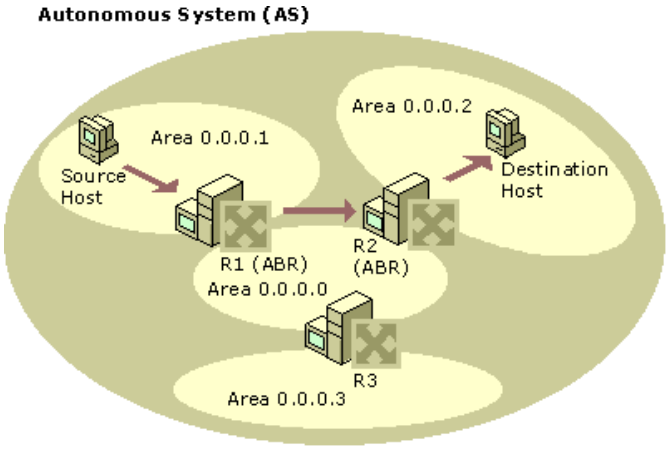


Figure 3.19 Inter-Area Routing in OSPF

Note OSPF routers do not make routing decisions based on area IDs. All routing decisions are based on the entries in the IP routing table. For example, in the inter-area routing shown in Figure 3.19, the backbone routers are not explicitly forwarding the packet to Area 0.0.0.2. They are forwarding it along the least cost path to the route that is the best match for the destination IP address in the packet.

Virtual Links

It is possible to define areas in such a way that they do not have an ABR physically connected to the backbone. Backbone connectivity for the area is still possible by configuring a virtual link between the non-backbone area and the backbone.

Virtual links can be configured between any two routers that have an interface to a single common non-backbone area. The common non-backbone area is known as the transit area. The transit area must have an ABR that is connected to the backbone. Virtual links cannot be configured across multiple transit areas.

A virtual link is not a physical link. It is a logical link using the least cost path between the ABR of the non-backbone connected area and the backbone ABR of the transit area. A virtual adjacency across the virtual link is formed, and routing information is exchanged. Just as in physical adjacencies, the settings of the two virtual link routers (such as the password, the Hello Interval, and Dead Interval) must match before an adjacency can be successfully established.

In Figure 3.20, Area 0.0.0.3 does not have a router physically connected to the backbone, Area 0.0.0.0. Therefore, a virtual link is configured across the transit area of Area 0.0.0.2 between routers R2 and R3. R2 and R3 are known as virtual link neighbors.

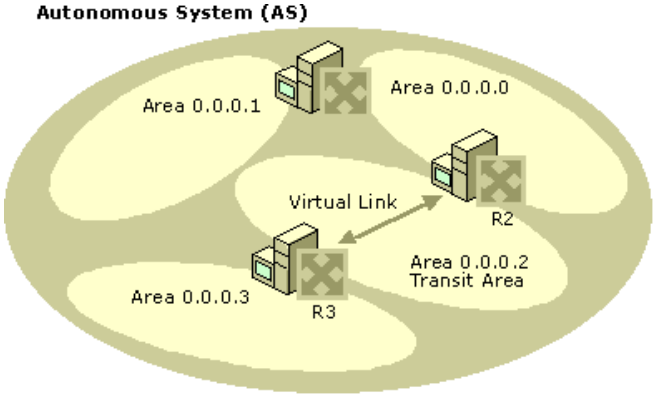


Figure 3.20 OSPF Virtual Link

Configuring Virtual Links

To configure a virtual link, configure a virtual interface in the **Routing and Remote Access** snap-in from the properties of the **OSPF** routing protocol on each virtual link neighbor with the following:

- The Area ID of the transit area for the virtual link.
- The OSPF Router ID of the virtual link neighbor.
- Adjacency settings such as the Retransmit Interval, Hello Interval, Dead Interval, and Password. The Hello Interval, Dead Interval, and Password must match between the two routers on each side of the virtual link in order for an adjacency to be established. The Retransmit Interval specifies how long the OSPF router waits before retransmitting Link State packets.

To view the virtual links

- In the **Routing and Remote Access** snap-in, in the **IP Routing** container, right-click **OSPF**, and then click **Show Virtual Interfaces**.

  The **Virtual Interfaces** window displays all configured virtual interfaces and their state.

External Routes

An external route is defined as any route that is not within the OSPF AS. External routes can come from many sources:

- Other routing protocols such as RIP for IP (v1 and v2), EGP, or BGP.
- Static routes.
- Routes set on the router through SNMP.

External routes are learned and propagated throughout the OSPF AS through one or more ASBRs. The ASBR advertises the availability of external routes using a series of external route LSAs. The external route LSAs are flooded throughout the AS (except in stub areas) and become part of the SPF Tree and routing table calculation. Traffic to external networks is routed within the AS using the least cost path to the ASBR.

Figure 3.21 shows an AS with an ASBR and external routes.

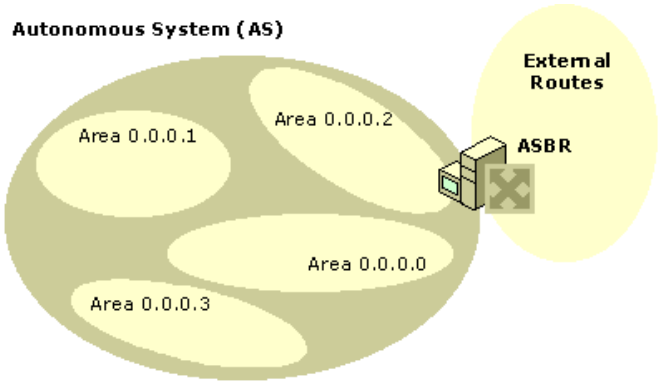


Figure 3.21 OSPF External Routes

External Route Filters

By default, OSPF routers acting as ASBRs import and advertise all external routes. You might want to filter out external routes to protect the AS from improper or malicious routing information.

For the Windows 2000 Router, external routes can be filtered on the ASBR by the external route source or by the individual route. You can configure the ASBR to accept or ignore the routes of certain external sources such as routing protocols (RIP v2) or other sources (static routes or SNMP). You can also configure the ASBR to accept or discard specific routes by configuring one or multiple [Destination, Network Mask] pairs.

A combination of these filters configured at the ASBR can ensure that the OSPF AS only receives the correct external routes from the proper sources.

ASBRs and Default Routes

When a router is configured to be an ASBR, it advertises by default all external routes including its own default static route. This default route needs to be valid for all OSPF routers in your AS. An example of an invalid default route is one that points to another router within the OSPF AS. The router used as the default gateway ends up with a default route with the next hop IP address of itself. If this occurs, the packets forwarded using the default route on that router are dropped.

If the default route is not valid for all OSPF routers, it should not be advertised. A valid default route would have the next hop gateway address external to your OSPF AS. This route would only be configured on the router that can directly reach the external network.

There are two ways to avoid this problem:

1. Do not configure a default gateway on the ASBR.
2. If the ASBR must have a default gateway, create an OSPF external route filter to filter out its own default route (destination of 0.0.0.0 with a network mask of 0.0.0.0).

Stub Areas

To further reduce the amount of routing information flooded into areas, OSPF allows the use of stub areas. A stub area can contain a single entry and exit point (a single ABR), or multiple ABRs when any of the ABRs can be used to reach external route destinations. For stub areas with multiple ABRs, the external routes are advertised by an ASBR that is outside of the area. AS external routes are not flooded into and throughout a stub area. Routing to all AS external networks in a stub area is done through a default route (destination 0.0.0.0 with the network mask of 0.0.0.0). Thus, a single entry in the routing tables of the routers in a stub area is used to route to all AS external locations.

To create the default route, the ABR of a stub area advertises a default route into the stub area. The default route is flooded to all the routers within the stub area, but not outside the stub area. The default route is used by the routers in a stub area for any destination IP address that is not reachable within the AS.

For example, Area 0.0.0.3 in Figure 3.22 is configured as a stub area because all external traffic must travel through its single ABR, router R3. R3 advertises a default route for distribution inside Area 0.0.0.3 instead of flooding the AS external networks into the area.

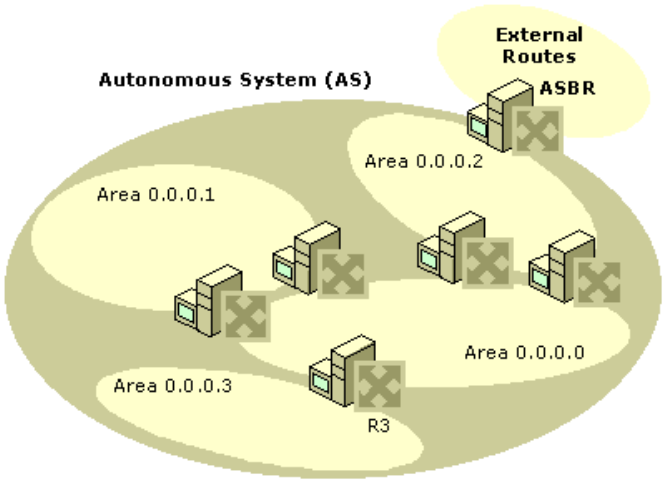


Figure 3.22 OSPF Stub Areas

All routers in a stub area must be configured so that they do not import or flood AS external routes within the stub area. Therefore, all area configurations for all router interfaces within a stub area must be configured for a stub area. Whether or not a router interface is in a stub area is indicated in a special option bit called the E-bit in the OSPF Hello packet. When the E-bit is set to 1, the router is allowed to accept and flood AS external routes. When the E-bit is set to 0, the router is not allowed to accept and flood AS external routes. Routers receiving Hello packets on interfaces verify that the E-bit of the received Hello packet matches their configuration before establishing an adjacency.

Stub areas have the following constraints:

- Virtual links cannot be configured using a stub area as a transit area.
- An ASBR cannot be placed inside a stub area.

Stub areas as defined in the OSPF RFC collapse all external routes into a single default route. Therefore, within a stub area, a router's routing table contains intra-area routes, inter-area routes, and a default route. The Windows 2000 Router also supports the collapsing of all non-intra-area routes into a single default route. This is known as a totally stubby area. A router's routing table within a totally stubby area contains intra-area routes and a default route. The default route summarizes all inter-area routes and all external routes.

To configure a stub area on a Windows 2000 Router, when configuring general properties for an area, select the **Stub area** check box and select the **Import summary advertisements** check box. To configure a totally stubby area on a Windows 2000 Router, select the **Stub area** check box but do not select the **Import summary advertisements** check box.

Troubleshooting OSPF

If an OSPF environment is properly configured, OSPF routers learn all the least cost routes from adjacent OSPF routers after convergence. The exact list of routes added by OSPF to the IP routing table depends, among other factors, on whether or not areas are configured to summarize their routes, whether or not stub areas or totally stubby areas are being used, and whether or not ASBRs and route filtering is being used.

Most OSPF problems are related to the formation of adjacencies, either physical or logical (through virtual links). If adjacencies cannot form, the LSDBs cannot be synchronized and the OSPF routes will not accurately reflect the topology of the internetwork. Other OSPF problems are related to the lack of routes or existence of improper routes in the IP routing table.

Adjacency Is Not Forming

- Before proceeding, verify that the two neighboring routers *should* form an adjacency. If the two routers are the only routers on the network, an adjacency should form. If there are more than two routers on the network, adjacencies only form with the DR and BDR. If the two routers have already formed adjacencies with the DR and the BDR, they will not form adjacencies with each other. In this case, their neighbor should appear as **2-way** under neighbor state.
- Ping the neighboring router to ensure basic IP and network connectivity. Use the **tracert** command to trace the route to the neighboring router. There should not be any routers between the neighboring routers.
- Use OSPF logging to log errors and warnings to record information about why the adjacency is not forming. To obtain additional information about OSPF processes, enable tracing for the OSPF component. For more information about tracing, see "Routing and Remote Access Service" in this book.
- Verify that the areas are enabled for authentication and the OSPF interfaces are using the same password. Windows 2000 OSPF routers have authentication enabled by default and the default password is "12345678". Change the authentication to match all neighboring OSPF routers on the same network. The password can vary per network.
- Verify that the routers are configured for the same Hello Interval and Dead Interval. By default the Hello Interval is 10 seconds and the Dead Interval is 40 seconds.
- Verify that the routers agree as to whether the area to which the common network belongs is a stub area or not.
- Verify that the interfaces of the neighboring routers are configured with the same Area ID.
- If the routers are on a Non-Broadcast Multiple Access (NBMA) network such as X.25 or Frame Relay and the connection to the NBMA network appears as a single adapter (rather than separate adapters for each virtual circuit), their neighbors must be manually configured using the unicast IP address of the neighbor or neighbors to which the link state information needs to be sent. Also verify that Router Priorities are configured so that one router can become the DR for the network.
- On broadcast networks (Ethernet, Token Ring, FDDI) or NBMA networks (X.25, Frame Relay), verify that all routers do not have a Router Priority of 0. At least one router must have a Router Priority of 1 or greater so that it can become the DR for the network.
- Verify that IP packet filtering is not preventing the receiving (through input filters) or sending (through output filters) of OSPF messages on the router interfaces enabled for OSPF. OSPF uses the IP protocol number 89.
- Verify that TCP/IP packet filtering is not preventing the receiving of OSPF messages on the interfaces enabled for OSPF.

Virtual Link Is Not Forming

- Verify that the virtual link neighbor routers are configured for the same password, Hello Interval, and Dead Interval.
- For each router, verify that the virtual link neighbor's Router ID is correctly configured.
- Verify that both virtual link neighbors are configured for the correct transit area ID.
- For large internetworks with substantial round-trip delays across the transit area, verify that the re-transmit interval is long enough.

Lack of OSPF Routes or Existence of Improper OSPF Routes

- If you are not receiving summarized OSPF routes for an area, verify that all the ABRs for the area are configured with the proper {Destination, Network Mask} pairs summarizing that area's routes.
- If you are receiving both individual and summarized OSPF routes for an area, verify that all the ABRs for the area are configured with the proper {Destination, Network Mask} pairs summarizing that area's routes.
- If you are not receiving external routes from the ASBR, verify that the source and route filtering configured on the ASBR is not too restrictive, preventing proper routes from being propagated to the OSPF AS. External source and route filtering is configured on the **External Routing** tab on the properties of the **OSPF** routing protocol in the **Routing and Remote Access** snap-in.
- Verify that all ABRs are either physically connected to the backbone or logically connected to the backbone using a virtual link. There should not be backdoor routers—routers connecting two areas without going through the backbone.

DHCP Relay Agent

The Windows 2000 Router is an RFC 1542-compliant Boot Protocol (BOOTP) relay agent relaying Dynamic Host Configuration Protocol (DHCP) messages between DHCP clients and DHCP servers on different IP networks. In this role, the Windows 2000 Router functions as the DHCP Relay Agent. Without a DHCP Relay Agent, a DHCP server is required on every subnet that contains DHCP clients. The DHCP Relay Agent takes broadcasted DHCP messages from DHCP clients and forwards them to the IP addresses of DHCP servers. The responses from the DHCP server are sent to the IP address of the DHCP Relay Agent, which then forwards them to the DHCP client.

For more information about DHCP and its implementation in Windows 2000, see "Dynamic Host Configuration Protocol" in the *Microsoft® Windows® 2000 Server Resource Kit TCP/IP Core Networking Guide*.

DHCP Across IP Routers

In a large IP internetwork, DHCP servers should be placed in strategic locations servicing DHCP clients of multiple networks. For this configuration to work effectively, DHCP messages must be able to cross IP routers using a DHCP Relay Agent.

In addition to propagating DHCP messages, a DHCP Relay Agent takes an active role in recording information necessary for DHCP configuration and helps direct DHCP messages between the DHCP server and the DHCP client.

Initial DHCP Configuration

Initial DHCP configuration is done by a DHCP client that has never leased an IP address, has released its IP address, or has received a DHCPNack in response to attempting to lease a previous IP address. The initial DHCP configuration process consists of four DHCP messages: DHCPDiscover, DHCPOffer, DHCPRequest, DHCPAck.

DHCPDiscover

The DHCP client sends the DHCPDiscover, containing the MAC address of the DHCP client, to the limited broadcast IP address (255.255.255.255) and the MAC-level broadcast address. The DHCP Relay Agent receives and processes the DHCPDiscover.

As established in RFC 1542, the DHCP Relay Agent can forward the packet to an IP broadcast, multicast, or unicast address. In practice, DHCP Relay Agents forward DHCPDiscover messages to unicast IP addresses which correspond to DHCP servers. Before forwarding the original DHCPDiscover message, the DHCP Relay Agent makes the following changes:

- Increments the Hop Count field in the DHCP header. The DHCP Hop Count field is separate from the Time to Live (TTL) field in the IP header and is used to indicate on how many networks this DHCPDiscover has existed as a broadcast. When the configured maximum Hop Count is exceeded, the DHCPDiscover is silently discarded. The default maximum hop count for the Windows 2000 DHCP Relay Agent is 4.
- If needed, updates the Relay IP Address field (also known as the Gateway IP Address field) in the DHCP header. When the DHCP client sends the DHCPDiscover message, the Relay IP Address field is set to 0.0.0.0. If the Relay IP Address is 0.0.0.0, the DHCP Relay Agent records the IP address of the interface on which the DHCPDiscover message was received. If the Relay IP Address is not 0.0.0.0, the DHCP Relay Agent does not modify it. The Relay IP Address field records the first router interface encountered by the DHCPDiscover message.
- Changes the source IP address of the DHCPDiscover message to the IP address of the interface on which the broadcasted DHCPDiscover was received.
- Changes the destination IP address of the DHCPDiscover message to the configured unicast address of the DHCP server.

The DHCP Relay Agent sends the DHCPDiscover message as a unicasted IP packet rather than as an IP and MAC-level broadcast. If the DHCP Relay Agent is configured with multiple DHCP servers, it sends each DHCP server a copy of the DHCPDiscover message.

DHCPOffer

When responding to the DHCP client's request for an IP address, the DHCP server uses the Relay IP Address field in the following ways:

- The Relay IP Address and the subnet masks of the server's configured scopes are compared through a logical **AND** comparison to find a scope whose network ID matches the network ID of the Relay IP Address. When a match is found, the DHCP server allocates an IP address from that scope.
- When sending the offer back to the client, the DHCP server sends the DHCPOffer message to the Relay IP Address as the destination IP address.

Once received by the DHCP Relay Agent, the Relay IP Address is used to determine which interface to which the DHCPOffer message is to be forwarded. It then forwards the DHCPOffer message to the client using the offered IP address as the destination IP address and the client's MAC address as the destination MAC address.

DHCPRequest

As it does with the DHCPDiscover message, the DHCP client sends the DHCPRequest message, containing the MAC address of the client, to the limited IP broadcast address 255.255.255.255 and to the MAC-level broadcast address. The DHCP Relay Agent receives this packet and forwards it as a directed IP packet to the configured DHCP server or servers.

DHCPAck

The DHCP server initially sends the DHCPAck message to the Relay IP Address, as it did with the DHCPOffer message. When the DHCP Relay Agent receives the DHCPAck message, it re-addresses it to the client's offered IP address and MAC address.

Rebooted Renewal

When a Microsoft-based DHCP client shuts down, it does not send a DHCPRELEASE message to the DHCP server. Instead, when the DHCP client restarts, it attempts to obtain the IP address it was last using through a DHCPRequest and DHCPAck exchange of messages.

DHCPRequest

When a Microsoft-based DHCP client reboots, it attempts to lease its previously allocated IP address through a broadcasted DHCPRequest message. The DHCPRequest, sent to the limited IP broadcast address 255.255.255.255 and to the MAC-level broadcast address, contains the MAC address and the previously allocated IP address of the client. The DHCP Relay Agent receives this packet and treats the message in much the same way as a DHCPDiscover message. Before forwarding, the DHCP Relay Agent:

- Increments the Hop Count field in the DHCP header.
- Records the IP address of the interface on which the DHCPRequest message was received in the Relay IP Address field.
- Changes the source IP address to the IP address of the interface on which the broadcasted DHCPDiscover message was received.
- Changes the destination IP address to the unicast address of the DHCP server recorded in the DHCPRequest and forwards it as a directed IP packet.

DHCPAck and DHCPNack

When the DHCP server receives the DHCPRequest, it compares the network ID of client's previously allocated IP address to the network ID of the Relay IP Address.

- If the two network IDs are the same and the IP address can be reallocated to the DHCP client, the DHCP server initially sends a DHCPAck to the IP address found in the Relay IP Address field. When the DHCP Relay Agent receives the DHCPAck, it re-addresses it to the client's offered IP address and MAC address.
- If the two network IDs are the same and the IP address cannot be reallocated to the DHCP client, the DHCP server initially sends a DHCPNack to the IP address found in the Relay IP Address field. When the DHCP Relay Agent receives the DHCPNack, it re-addresses it to the client's offered IP address and MAC address. At this point, the DHCP client must restart the IP address allocation process with a DHCPDiscover.
- If the two network IDs are not the same, the DHCP client has moved to a different subnet, and the DHCP server sends a DHCPNack

to the IP address found in the Relay IP Address field. When the DHCP Relay Agent receives the DHCPNack, it re-addresses it to the client's offered IP address and MAC address. At this point, the DHCP client must restart the IP address allocation process with a DHCPDiscover.

Troubleshooting the DHCP Relay Agent

If the Windows 2000 DHCP Relay Agent is not providing relay services for DHCP clients on a network, check for the following:

- Verify that the interface on the Windows 2000 Router that connects to the network where the DHCP clients are located is added to the DHCP Relay Agent IP routing protocol.
- Verify that the **Relay DHCP packets** check box is selected for the DHCP Relay Agent interface connected to the network where the DHCP clients are located.
- Verify that the IP addresses of DHCP servers configured on the global properties of the DHCP Relay Agent are the correct IP addresses for DHCP servers on your internetwork.
- From the router with the DHCP Relay Agent enabled, use the PING utility to ping each of the DHCP servers configured in the global DHCP Relay Agent dialog. If you cannot ping the DHCP servers from the DHCP Relay Agent router, troubleshoot the lack of connectivity between the DHCP Relay Agent router and the DHCP server or servers.
- Verify that IP packet filtering is not preventing the receiving (through input filters) or sending (through output filters) of DHCP traffic. DHCP traffic uses the UDP ports of 67 and 68.
- Verify that TCP/IP filtering on the router interfaces is not preventing the receiving of DHCP traffic.

Network Address Translator

A Network Address Translator (NAT) is an IP router defined in RFC 1631 that can translate IP addresses and TCP/UDP port numbers of packets as they are being forwarded. Consider a small business network with multiple computers connecting to the Internet. A small business would normally have to obtain an Internet Service Provider (ISP)-allocated public IP address for each computer on their network. With the NAT, however, the small business can use private addressing (as described in RFC 1597) and have the NAT map its private addresses to a single or to multiple public IP addresses as allocated by its ISP.

For example, if a small business is using the 10.0.0.0 private network for its intranet and has been granted the public IP address of 198.200.200.1 by its ISP, the NAT maps (using static or dynamic mappings) all private IP addresses being used on network 10.0.0.0 to the public IP address of 198.200.200.1.

When a private user on the small business intranet connects to an Internet resource, the user's IP stack creates an IP packet with the following values set in the IP and TCP or UDP headers (bold text indicates the fields changed by the NAT):

- Destination IP Address: Internet resource IP address
- Source IP Address: **Private IP address**
- Destination Port: Internet resource TCP or UDP port
- Source Port: **Source application TCP or UDP port**

The source host or another router forwards this IP packet to the NAT, which translates the addresses of the outgoing packet as follows (bold text indicates the fields changed by the NAT):

- Destination IP Address: Internet resource IP address
- Source IP Address: **ISP-allocated public address**
- Destination Port: Internet resource TCP or UDP port
- Source Port: **Remapped source application TCP or UDP port**

The NAT sends the remapped IP packet over the Internet. The responding computer sends back the response to the NAT. When received by the NAT, the packet contains the following addressing information (bold text indicates the fields changed by the NAT):

- Destination IP Address: **ISP-allocated public address**
- Source IP Address: Internet resource IP address
- Destination Port: **Remapped source application TCP or UDP port**
- Source Port: Internet resource TCP or UDP port

When the NAT maps and translates the addresses and forwards the packet to the intranet client, it contains the following addressing information (bold text indicates the fields changed by the NAT):

- Destination IP Address: **Private IP address**
- Source IP Address: Internet resource IP address
- Destination Port: **Source application TCP or UDP port**
- Source Port: Internet resource TCP or UDP port

For outgoing packets, the source IP address and TCP/UDP port numbers are mapped to a public source IP address and a possibly changed TCP/UDP port number. For incoming packets, the destination IP address and TCP/UDP port numbers are mapped to the private IP address and original TCP/UDP port number.

Static and Dynamic Address Mapping

The NAT can use either static or dynamic mapping. A static mapping is configured so that traffic is always mapped a specific way. You could map all traffic to and from a specific private network location to a specific Internet location. For instance, to set up a Web server on a computer on your private network, you create a static mapping that maps [Public IP Address, TCP Port 80] to [Private IP Address, TCP Port 80].

Dynamic mappings are created when users on the private network initiate traffic with Internet locations. The NAT automatically adds these mappings to its mapping table and refreshes them with each use. Dynamic mappings that are not refreshed are removed from the NAT mapping table after a configurable amount of time. For TCP connections, the default time-out is 24 hours. For UDP traffic, the default time-out is one minute.

Proper Translation of Header Fields

By default, a NAT translates IP addresses and TCP/UDP ports. These modifications to the IP datagram require the modification and recalculation of the following fields in the IP, TCP, and UDP headers:

- Source IP Address (outbound from private network), Destination IP Address (inbound to private network)
- IP Checksum
- Source Port (outbound from private network), Destination Port (inbound to private network)
- TCP Checksum

- UDP Checksum

If the IP address and port information is only in the IP and TCP/UDP headers—for example, with HTTP (or World Wide Web) traffic, the application protocol can be translated transparently. There are applications and protocols, however, that carry IP or port addressing information within their headers. FTP, for example, stores the dotted decimal representation of IP addresses in the FTP header for the FTP **port** command. If the NAT does not properly translate the IP address, connectivity problems can occur. Additionally, in the case of FTP, because the IP address is stored in dotted decimal format, the translated IP address in the FTP header can be a different size. Therefore, the NAT must also modify TCP sequence numbers to ensure that no data is lost.

NAT Editors

In the case where the NAT component must additionally translate and adjust the payload beyond the IP, TCP, and UDP headers, a NAT editor is required. A NAT editor is an installable component that can properly modify otherwise nontranslatable payloads so that they can be forwarded across a NAT.

Windows 2000 includes built-in NAT editors for the following protocols:

- FTP
- ICMP
- PPTP
- NetBIOS over TCP/IP

Additionally, the NAT routing protocol includes proxy software for the following protocols:

- Direct Play
- LDAP-based ILS registration
- RPC

Note IPSec traffic is not translatable.

NAT Processes in the Windows 2000 Router

For the Windows 2000 Routing and Remote Access service, the NAT component is a routing protocol known as Network Address Translation or NAT. The NAT component can either be enabled by adding Network Address Translation as a routing protocol in the Routing and Remote Access snap-in.

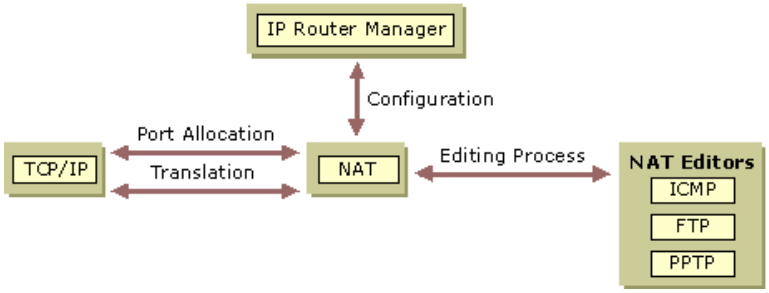
Note NAT services are also available with the Internet connection sharing feature available from the Network and Dial-up Connections folder. Internet connection sharing performs the same function as the NAT routing protocol in the Routing and Remote Access service but it allows very little configuration flexibility. For information about configuring Internet connection sharing and why you would choose Internet connection sharing over the NAT routing protocol of the Routing and Remote Access service, see Windows 2000 Server Help.

Installed with the NAT routing protocol are a series of NAT editors. NAT consults the editors when the payload of the packet being translated matches one of the installed editors. The editors modify the payload and return the result to the NAT component.

NAT interacts with the TCP/IP protocol in two important ways:

- To support dynamic port mappings, the NAT component requests unique TCP and UDP port numbers from the TCP/IP protocol stack when needed.
- With TCP/IP so that packets being sent between the private network and the Internet are first passed to the NAT component for translation.

Figure 3.23 shows the NAT components and their relation to TCP/IP and other router components.



If your browser does not support inline frames, click here to view on a separate page.

Figure 3.23 NAT Components

Outbound Internet Traffic

For traffic from the private network that is outbound on the Internet interface, the NAT first assesses whether or not an address/port mapping, static or dynamic, exists for the packet. If not, a dynamic mapping is created. The NAT creates a mapping depending on whether there are single or multiple public IP addresses available.

- If a single public IP address is available, the NAT requests a new unique TCP or UDP port for the public IP address and uses that as the mapped port.
- If multiple public IP addresses are available, the NAT performs private IP address to public IP address mapping. For these mappings, the ports are not translated. When the last public IP address is needed, the NAT switches to performing address and port mapping as it would in the case of the single public IP address.

After mapping, the NAT checks for editors and invokes one if necessary. After editing, the NAT modifies the TCP, UDP, and IP headers and forwards the frame using the Internet interface.

Figure 3.24 shows the NAT processing for outbound Internet traffic.

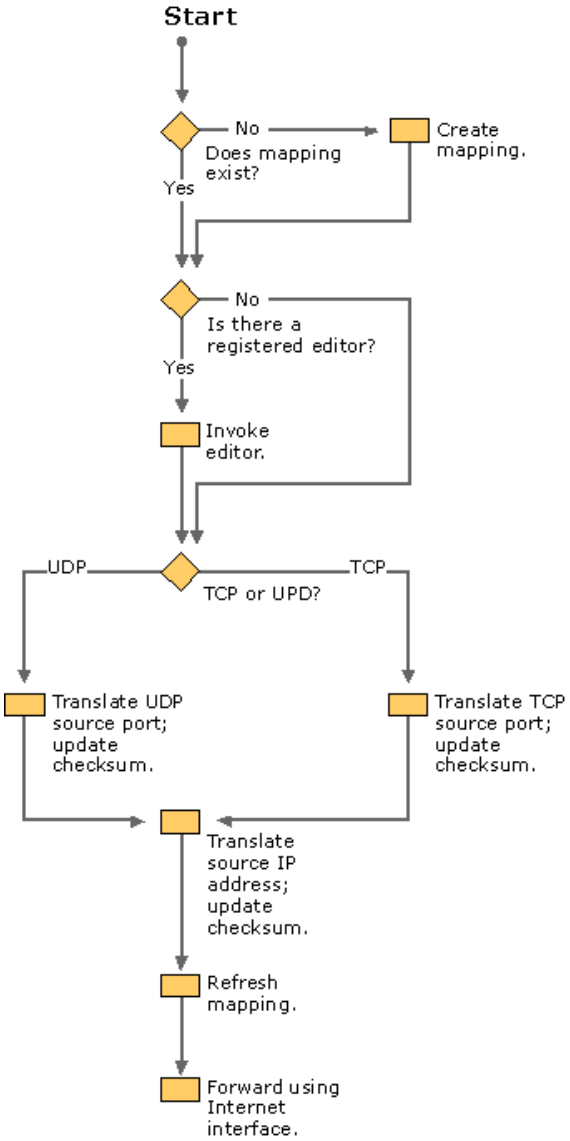


Figure 3.24 NAT Processing of Outbound Internet Traffic

Inbound Internet Traffic

For traffic from the private network that is inbound on the Internet interface, the NAT first assesses whether an address/port mapping, static or dynamic, exists for the packet. If a mapping does not exist for the packet, it is silently discarded by the NAT.

This behavior protects the private network from malicious users on the Internet. The only way that Internet traffic is forwarded to the private network is either in response to traffic initiated by a private network user that created a dynamic mapping or because a static mapping exists so that Internet users can access specific resources on the private network.

After mapping, the NAT checks for editors and invokes one if necessary. After editing, the NAT modifies the TCP, UDP, and IP headers and forwards the frame using the private network interface.

Figure 3.25 shows the NAT processing for inbound Internet traffic.

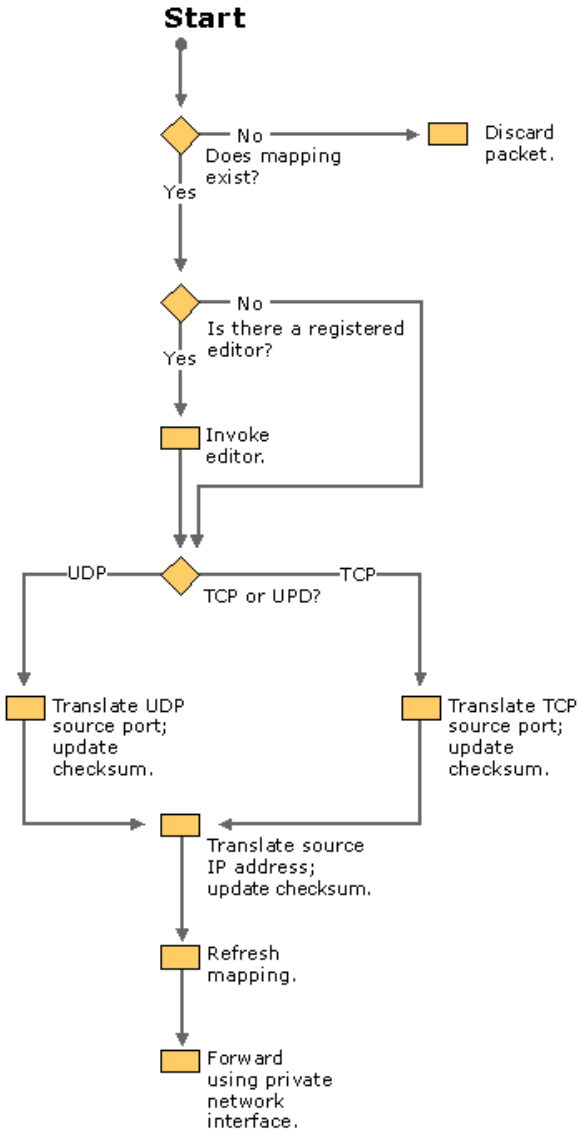


Figure 3.25 NAT Processing of Inbound Internet Traffic

Additional NAT Routing Protocol Components

To help simplify the configuration of small office/home office (SOHO) networks to the Internet, the NAT routing protocol for Windows 2000 also includes a DHCP allocator and a DNS proxy.

DHCP Allocator

The DHCP allocator component provides IP address configuration information to the other computers on the SOHO network. The DHCP allocator is a simplified DHCP server that allocates an IP address, a subnet mask, a default gateway, the IP address of a DNS server, and the IP address of a DNS server. You must configure computers on the DHCP network as DHCP clients in order to receive the IP configuration automatically. The default TCP/IP configuration for Windows 2000, Windows NT, Windows 95, and Windows 98 computers is as a DHCP client.

Table 3.5 lists the DHCP options in the DHCPOffer and DHCPAck messages issued by the DHCP allocator during the DHCP lease configuration process. You cannot modify these options or configure additional DHCP options.

Table 3.5 DHCP Allocator DHCP Options

Option Number	Option Value	Description
1	255.255.0.0	Subnet Mask
3	IP address of private interface	Router (default gateway)
6	IP address of private interface	DNS server (only issued if DNS proxy is enabled)
58 (0x3A)	5 minutes	Renewal time
59 (0x3B)	5 days	Rebinding time
51	7 days	IP address lease time
15 (0x0F)	primary domain name of NAT computer	DNS domain

The DHCP allocator only supports a single scope of IP addresses as configured from the **Address Assignment** tab on the properties of the **Network Address Translation (NAT)** routing protocol in the **Routing and Remote Access** snap-in. The DHCP allocator does not support multiple scopes, superscopes, or multicast scopes. If you need this functionality, you should install a DHCP server and disable the DHCP allocator component of the NAT routing protocol.

DNS Proxy

The DNS proxy component acts as a DNS server to the computers on the SOHO network. DNS queries sent by a SOHO computer to the NAT computer are re-sent by the NAT computer as DNS queries from the NAT computer to the NAT computer's configured DNS server. Responses to DNS queries corresponding to outstanding requests of SOHO computers received by the NAT computer are re-sent by the NAT computer to the original SOHO computer.

Troubleshooting NAT

Most NAT problems deal with the inability of the NAT to translate packets. Other problems are related to address allocation and name resolution.

The network address translation computer is not properly translating packets

- Verify that the interface on the Windows 2000 Router that connects to the Internet is added to the **Network Address Translation (NAT)** routing protocol.
- Verify that the **Public interface connected to the Internet** option on the **General** tab on the properties page of the Internet interface is selected.
- Verify that the **Private interface connected to private network** option on the **General** tab on the properties page of the private network interface is selected.
- If you only have a single public IP address, verify that the **Translate TCP/UDP headers** option on the **General** tab on the properties page of the Internet interface is selected.
- If you have multiple public IP addresses, verify that they are typed correctly in text boxes provided on the **Address Pool** tab on the properties page of the Internet interface. If your address pool includes an IP address that was not allocated to you by your ISP, inbound Internet traffic that is mapped to that IP address is routed by the ISP to another location.
- If you have some applications that do not seem to work through the NAT, try running them from the NAT computer. If they work from the NAT computer and not from a computer on the private network, the payload of the application might not be translatable. Check the protocol being used by the application against the list of supported NAT editors. If needed, contact the vendor of the application for information about how their application works in translated environments.
- Verify that IP packet filtering on the private network and Internet interfaces is not preventing the receiving (through input filters) or sending (through output filters) of Internet-based traffic.
- Verify that TCP/IP filtering on the private network and Internet interfaces is not preventing the receiving of traffic.
- For special ports, verify the configuration of the public address and port and the private address and port.

Private network hosts are not receiving IP address configuration

- Verify that the DHCP allocator is enabled from the **Address Assignment** tab of the properties of the **Network Address Translation (NAT)** routing protocol.

Name resolution for private network hosts is not working

- Verify that the DNS proxy is enabled from the **Address Assignment** tab of the properties of the **Network Address Translation (NAT)** routing protocol.
- Verify the name resolution configuration of the network address translation computer by using the ipconfig command. There are two ways that you can configure name resolution when dialing an ISP:
  - Statically assigned name servers

    You must manually configure the TCP/IP protocol with the IP address (or addresses) of the name servers provided by the ISP. If you have statically assigned name servers, you can use the ipconfig command at any time to get the IP addresses of your configured name servers.
  - Dynamically assigned name servers

    Manual configuration is not required. The IP addresses of the name servers provided by the ISP are dynamically assigned whenever you dial the ISP. If you have dynamically assigned name servers, you must run the ipconfig command after a connection to the ISP has been made.

IP Packet Filtering

To provide security, an IP router can allow or disallow the flow of very specific types of IP traffic. This capability, called IP packet filtering, provides a way for the network administrator to precisely define what IP traffic is received and sent by the router. IP packet filtering is an important element of connecting corporate intranets to public networks like the Internet.

IP packet filtering consists of creating a series of definitions called filters, which define for the router what types of traffic are allowed or disallowed on each interface. Filters can be set for incoming and outgoing traffic.

- Input filters define what inbound traffic on that interface the router is allowed to route or process.
- Output filters define what traffic the router is allowed to send from that interface.

Because you can configure both input and output filters for each interface, it is possible to create contradictory filters. For example, the input filter on one interface allows the inbound traffic but the output filter on the other interface does not allow the same traffic to be sent. The end result is that the traffic is not passed across the Windows 2000 Router.

Packet filtering can also be implemented on a non-router computer running Windows 2000 to filter incoming and outgoing traffic to a specific subset of traffic.

Packet filters should be implemented carefully to prevent the filters from being too restrictive, which would impair the functionality of other protocols that might be operating on the computer. For example, if a computer running Windows 2000 is also running Internet Information Services (IIS) as a Web server and packet filters are defined so that only Web-based traffic is allowed, you can not use PING (which uses ICMP Echo Requests and Echo Replies) to perform basic IP troubleshooting. If the Web server is a Silent RIP host, the filters prevent the Silent RIP process from receiving the RIP announcements.

Note When troubleshooting connectivity or IP-based network problems on a computer running Windows 2000 that is using packet filtering, first verify whether the packet filtering configured on that computer is preventing outgoing or incoming packets for the protocol having the problem.

Windows 2000 IP Packet Filtering

Windows 2000 IP packet filtering is based on exceptions. You can configure Windows 2000 to either pass all traffic except those disallowed by filters or to discard all traffic except those allowed by filters. For example, you might want to configure a filter to allow all traffic except Telnet traffic (TCP port 23). Or you might want to set up filters on a dedicated Web server to process only Web-based TCP traffic (TCP port 80).

Note The Windows 2000 Router does not allow the use of user-definable filters where a network administrator can create a filter based on any field of the IP, TCP, UDP or ICMP header. The Windows 2000 Router does not support filtering on any protocols other than IP, TCP, UDP, and ICMP.

Windows 2000 allows filtering on various fields in IP, TCP, UDP, and ICMP headers of incoming and outgoing packets.

IP Header

In the IP header, filters can be defined for the following fields:

IP Protocol An identifier used to demultiplex the payload of an IP packet to an upper layer protocol. For example, TCP uses a Protocol of 6, UDP uses a Protocol of 17, and ICMP uses a Protocol of 1. When you select a protocol such as TCP, UDP, or ICMP in the **IP Filters** dialog box, the default values for those protocols are assumed. Windows 2000 allows you to type any value in the **IP Protocol** text box.

Source IP Address The IP address of the source host, which can be configured with a subnet mask, allowing an entire range of IP addresses (corresponding to an IP network) to be specified with a single filter entry.

Destination IP Address The IP address of the destination host which can be configured with a subnet mask, allowing an entire range of IP addresses (corresponding to an IP network) to be specified with a single filter entry.

TCP Header

In the TCP header, filters can be defined for two fields: the TCP Source Port field, used to identify the source process which is sending the TCP segment; and for the TCP Destination Port, used to identify the destination process for the TCP segment.

Note The Windows 2000 Router does not support the configuration of a range of TCP ports. For a range of TCP ports, a separate filter for each port in the range must be configured.

UDP Header

In the UDP header, filters can be defined for two fields: the UDP Source Port field, used to identify the source process which is sending the UDP message; and for the UDP Destination Port, used to identify the destination process for the UDP message.

Note The Windows 2000 Router does not support the configuration of a range of UDP ports. For a range of UDP ports, a separate filter for each port in the range must be configured.

ICMP Header

In the ICMP header, filters can be defined for two fields: the ICMP Type field, indicating the type of ICMP packet (such as Echo Request or Echo Reply); and for the ICMP Code field, indicating one of the possible multiple functions within a specified type. If there is only one function within a type, the Code field is set to 0.

Table 3.6 lists commonly used ICMP types and codes.

Table 3.6 Common ICMP Types and Codes

ICMP Type	ICMP Code	Use
0	0	Echo Reply
8	0	Echo Request
3	0	Destination Unreachable - Network Unreachable
3	1	Destination Unreachable - Host Unreachable
3	2	Destination Unreachable - Protocol Unreachable
3	3	Destination Unreachable - Port Unreachable
3	4	Destination Unreachable - Fragmentation Needed and Don't Fragment Flag set
4	0	Source Quench
5	1	Redirect - Redirected datagrams for the host
9	0	Router Advertisement
10	0	Router Solicitation
11	0	Time Exceeded - TTL expiration
11	1	Time Exceeded - Fragmentation Reassembly expiration
12	0	Parameter Problem

Note For a complete list of ICMP types and codes, see the link at http://windows.microsoft.com/windows2000/reskit/webresources .

Input Filters

Input filters are configured on an exception basis. You can configure the filter action to either receive all traffic except that which is specified, or to drop all traffic except that which is specified.

When multiple filters are configured, the separate filters applied to the inbound packet are compared through a logical **OR**. If the packet matches at least one of the configured filters, it is received or dropped depending on the filter action setting.

Output Filters

Output filters are configured on an exception basis. You can configure the filter action to either transit all traffic except that which is specified, or to drop all traffic except that which is specified.

When multiple filters are configured, the separate filters applied to the outbound packet are compared through a logical **OR**. If the packet matches at least one of the configured filters, it is transmitted or dropped depending on the filter action setting.

Configuring a Filter

When adding or editing an input or an output filter, you configure the parameters of the filter in the **Add IP Filter** or **Edit IP Filter** dialog boxes. When multiple parameters are configured on a particular filter, as the filter is applied to the incoming packet, the parameters of the filter are compared through a logical **AND**. The fields in the packet must match all of the configured parameters of the filter to meet the criteria of the filter.

Note You cannot configure separate active filters for **Receive all packets except those that meet the criteria below** and **Drop all**

packets except those that meet the criteria below.

You can configure the following fields in the **Add IP Filter** or **Edit IP Filter** dialog boxes:

Source Network

- IP Address: Type the source IP network ID or a source IP address.
- Subnet Mask: Type the subnet mask corresponding to the source network ID or type 255.255.255.255 for a source IP address. The subnet mask bits must encompass all of the bits being used in the IP Address field. The IP address cannot be more specific than the subnet mask.

Destination Network

- IP Address: Type the destination IP network ID or a destination IP address.
- Subnet Mask: Type the subnet mask corresponding to the destination network ID, or type 255.255.255.255 for a destination IP address. The subnet mask bits must encompass all of the bits being used in the IP Address field. The IP address cannot be more specific than the subnet mask.

Protocol

- TCP (Protocol = 6): Select this option to reach text boxes in which you type a source TCP port and a destination TCP port. One or both can be specified. If nothing is specified in these text boxes, they default to 0, meaning any port.
- TCP [established] (Protocol = 6): Select this option when you want to define TCP traffic (source TCP port and destination TCP port) for TCP connections established by or with the router.
- UDP (Protocol = 17): Select this option to reach text boxes in which you type a source UDP port and a destination UDP port. One or both can be specified. If nothing is specified in these text boxes, they default to 0, meaning any port.
- ICMP (Protocol = 1): Select this option to reach text boxes in which you type an ICMP code and an ICMP type. One or both can be specified. If nothing is specified in these text boxes, they default to 255, meaning any code or any type.
- Any: Select this option to make *any* IP protocol value assumed.
- Other: Select this option to reach the text box in which you type any IP protocol.

Filtering Scenarios

This section illustrates filter configurations for commonly implemented filtering scenarios.

Caution If you combine any of the sample sets of filters, make sure that the desired subset of traffic is allowed and the desired level of security is maintained. For example, if you combine the local host filtering and Web traffic filtering, due to the way that the filters are applied (**AND** is used within a filter; **OR** is used between filters), all traffic destined for the host is allowed. The Web traffic input filter is essentially ignored.

Local Host Filtering

Use local host filtering to ensure that only traffic destined for the host is allowed to be processed. This disables the forwarding of packets on the interface on which local host filtering is enabled. Local host filtering is used when an intranet is connected to the Internet and direct routing of packets between the intranet and the Internet is not desired. In this scenario, local host filtering is configured on the Internet interface.

Configure the following filters on the Internet interface. With these filters configured, only traffic destined for this host or for all hosts on the host's network, or multicast traffic is allowed on the interface.

Using the **Drop all packets except those that meet the criteria below** filter action, create a series of input filters with the following attributes:

Destination IP Address of Host IP Address

- In the **Add IP Filter** dialog box, select the **Destination network** check box, and then type the IP address of the host and the subnet mask of 255.255.255.255 in the appropriate text boxes.

Destination IP Address of Subnet Broadcast

1. In the **Add IP Filter** dialog box, select the **Destination network** check box, and then type the IP address of the host's subnet broadcast IP address and the subnet mask of 255.255.255.255 in the appropriate text boxes.
2. To define the Subnet broadcast, set all the host bits to 1. For example, if a host is configured with an IP address of 172.16.5.98 with a subnet mask of 255.255.255.0 (a subnet of the private IP network 172.16.0.0), this filter would be filtering on the Destination IP address of 172.16.5.255.

Destination IP Address of All Subnets-Directed Broadcast

- In the **Add IP Filter** dialog box, select the **Destination network** check box, and then type the IP address of the host's all subnets-directed broadcast address and the subnet mask of 255.255.255.255 in the appropriate text boxes.

The all subnets-directed broadcast is class-based broadcast address where the host bits before subnetting are set to all 1. For the example host, this filter would be filtering on the Destination IP address of 172.16.255.255. The filter for the all subnets-directed broadcast is only necessary when subnetting.

Destination IP Address of the IP Limited Broadcast

- In the **Add IP Filter** dialog box, select the **Destination network** check box, and then type the IP address of 255.255.255.255 and the subnet mask of 255.255.255.255 in the appropriate text boxes.

The Limited Broadcast is the destination IP address of 255.255.255.255.

Destination IP Address for All Possible Multicast Traffic

- In the **Add IP Filter** dialog box, select the **Destination network** check box, and then type the IP address of 224.0.0.0 and the subnet mask of 240.0.0.0 in the appropriate text boxes. All possible inbound multicast traffic is allowed on the interface.

Note Local host filtering on an interface effectively disables unicast routing on that interface because the only unicast traffic allowed through the interface is destined for the host. Transit traffic is dropped.

Web Traffic Filtering

Web traffic filtering is done on hosts that are Web servers so that only Web-based traffic to and from the Web server service on the host is allowed to be processed. This is done to secure the Web server from malicious attacks on other services running on the Web

server. For a Web server connected to the Internet, Web traffic filtering is configured on the Internet interface.

Using the **Drop all packets except those that meet the criteria below** filter action, configure the following filters to confine the allowed traffic to packets to and from the Web server service:

- An input filter for the Destination IP Address of Web server and the TCP Destination Port 80.
- An output filter for the Source IP Address of Web server and the TCP Source Port 80.

If these filters are the only filters configured, the only traffic allowed through the interface is TCP traffic to and from the Web server service on the Windows 2000 Server-based computer.

Note The preceding example assumes the default port of the Web server is port 80. If you are using a port other than 80, substitute the appropriate port for port 80 in these filters.

FTP Traffic Filtering

FTP traffic filtering is done on hosts that are FTP servers so that only FTP-based traffic to and from the FTP server service on the host is allowed to be processed. This is done to secure the FTP server from malicious attacks on other services running on the FTP server. For a FTP server connected to the Internet, FTP traffic filtering is configured on the Internet interface.

Using the **Drop all packets except those that meet the criteria below** filter action, configure the following filters to confine the allowed traffic to packets to and from the FTP server service:

- Input filters for the Destination IP Address of FTP Server and the TCP Destination Port 21 (the FTP control port), and for the Destination IP Address of FTP Server and the TCP Destination Port 20 (the FTP data port).
- Output filters for the Source IP Address of FTP Server and the TCP Source Port 21 (the FTP control port), and for the Source IP Address of FTP Server and the TCP Source Port 20 (the FTP data port).

If these filters are the only filters configured, the only traffic allowed through the interface is TCP traffic to and from the FTP server service on the Windows 2000 Server-based computer.

Note The preceding example assumes the default ports, 20 and 21, of the FTP server. If you are using ports other than 20 and 21, substitute the appropriate ports for ports 20 and 21 in these filters.

PPTP Traffic Filtering

Point-to-Point Tunneling Protocol (PPTP) traffic filtering is done on hosts that are PPTP servers so that only PPTP-based traffic to and from the PPTP server service on the host is allowed to be processed. This is done to secure the PPTP server from malicious attacks on other services running on the PPTP server. For a PPTP server connected to the Internet, PPTP traffic filtering is configured on the Internet interface.

Using the **Drop all packets except those that meet the criteria below** filter action, configure the following filters to confine the allowed traffic to packets to and from the PPTP service running on the server:

- Input filters for the Destination IP Address of the PPTP server and TCP Destination Port 1723, and for the Destination IP Address of the PPTP server and IP Protocol 47 (Generic Routing Encapsulation [GRE]).
- Output filters for the Source IP Address of the PPTP server and the TCP Source Port 1723, and for the Source IP Address of the PPTP server and the IP Protocol 47 (GRE).

If the PPTP server is also to be used as a PPTP client to initiate tunneled connections to branch offices in a virtual private network (VPN) scenario, configure the following additional filters:

- An input filter for the Destination IP Address of the PPTP server and the TCP (established) Source Port 1723.
- An output filter for the Source IP Address of the PPTP server and the TCP (established) Destination Port 1723.

The TCP (established) filter is used to allow only traffic on the TCP connection that was established by the PPTP client. If TCP (established) is not used, a malicious Internet hacker can penetrate the PPTP server by sending packets from applications using TCP port 1723.

If these filters are the only filters configured, the only traffic allowed through the interface is TCP traffic and tunneled data (GRE traffic) to and from the PPTP server and PPTP client on the Windows 2000 Server-based computer.

For more information about PPTP, see "Virtual Private Networking" in this book.

L2TP Server Filtering

Layer Two Tunneling Protocol (L2TP) over IPSec traffic filtering is done on hosts that are L2TP servers so that only L2TP-based traffic to and from the L2TP server service on the host is allowed to be processed. This is done to secure the L2TP server from malicious attacks on other services running on the L2TP server. For a L2TP server connected to the Internet, L2TP traffic filtering is configured on the Internet interface.

Using the **Drop all packets except those that meet the criteria below** filter action, configure the following filters to confine the allowed traffic to packets to and from the server running L2TP:

- An input filter for the Destination IP Address of the L2TP server and UDP Destination Port 1701.
- An input filter for the Destination IP Address of the L2TP server and UDP Destination Port 500.
- An output filter for the Source IP Address of the L2TP server and the UDP Source Port 1701.
- An output filter for the Source IP Address of the L2TP server and the UDP Source Port 500.

The filters for UDP port 1701 are for the L2TP protocol. The filters for UDP port 500 are for the Internet Key Exchange (IKE) used to create the IPSec security association. Packet filters for the IPSec Encapsulating Security Payload (ESP) header using IP protocol 50 are not needed because the inbound and outbound packets are first processed by IPSec, which adds or removes the ESP header before the Routing and Remote Access service IP packet filters are applied.

If these filters are the only filters configured, the only traffic allowed through the interface is UDP traffic to and from the L2TP server and client on the Windows 2000 Server-based computer.

For more information about L2TP over IPSec, see "Virtual Private Networking" in this book.

Denying Spoofed Packets from Private IP Addresses

Another method of performing denial of service attacks is to flood servers with packets, such as TCP connection request packets, from addresses to which there can be no reply. In these cases, the malicious users spoof, or substitute, the source IP address of the packets with something other than the IP address of the interface on which the packets originated. An easy address to spoof is a private address because a response sent to a private address on the Internet results in an ICMP Destination Unreachable message.

To drop Internet traffic from spoofed private IP addresses, configure input filters on the Internet interface to accept all packets except the following:

- The Source IP Address of 10.0.0.0 with the subnet mask 255.0.0.0.
- The Source IP Address of 172.16.0.0 with the subnet mask 255.240.0.0.
- The Source IP Address of 192.168.0.0 with the subnet mask 255.255.0.0.

Fragmentation Filtering

The Routing and Remote Access service also supports the filtering of fragmented IP datagrams. A fragmented IP datagram is an IP datagram that contains a fragment of an IP payload. Source hosts or routers fragment IP payloads so that the resulting IP datagram is small enough to be sent on the network segment of the next hop. Routing and Remote Access service fragmentation filtering only applies to incoming traffic.

To enable fragmentation filtering

1. In the **Routing and Remote Access** snap-in, open the **IP Routing** container for the desired server.
2. Open the **General** container, right-click the desired interface, and click **Properties**.
3. From the **General** tab, select the **Enable fragmentation checking** check box.

To prevent the router from forwarding fragmented IP packets for transit traffic on any interface, select this setting on all interfaces of the router. This does not prevent the forwarding of fragmented packets sent from the router.

Fragmentation filtering can be employed to prevent the Ping of Death, a denial of service attack where malicious users send one or multiple 64-KB ICMP Echo Request messages. The 64-KB messages are fragmented and must be reassembled at the destination host. For each separate 64-KB message, the TCP/IP protocol must allocate memory, tables, timers, and other resources. With enough fragmented messages, a Windows 2000 Server-based computer can become bogged down so that the servicing of valid information requests is impaired. By using fragmentation filtering, incoming fragmented IP datagrams are immediately discarded.

ICMP Router Discovery

The Windows 2000 Router includes support for the router advertisement and discovery scheme as documented in RFC 1256. To aid in the ease of configuring IP hosts with the IP addresses of local routers and to provide a way for hosts to sense downed routers, RFC 1256 describes the use of ICMP messages to send Router Advertisements and Router Solicitations.

Router Advertisements

A router sends out a periodic Router Advertisement using an ICMP message (Type 9, Code 0). The Router Advertisement can be sent to the all-hosts IP multicast address of 224.0.0.1, the local IP broadcast address, or the limited broadcast address (255.255.255.255). In practice, the Router Advertisement is sent to the multicast address. Router Advertisements are explicit notifications to the hosts on the network that the router is still available.

The Router Advertisement message contains an Advertisement Lifetime, the time after the last received Router Advertisement that the router can be considered down (default of 30 minutes), and a Preference Level, an indication of the router's preferred status as the default gateway for the network. The highest preference level router becomes the default gateway.

Router Solicitations

When a host that supports RFC 1256 needs to be configured with a default gateway (either upon initialization or because its default gateway is down), it sends out a Router Solicitation using an ICMP message (Type 10, Code 0). The Router Solicitation can be sent to the all-routers IP multicast address of 224.0.0.2, the local IP broadcast address, or the limited broadcast address (255.255.255.255). In practice, hosts send Router Solicitation messages to the multicast address. Routers on the host's network that support RFC 1256 immediately respond with a Router Advertisement, and the host chooses the router with the highest preference level as its default gateway.

Note ICMP router discovery is not a routing protocol. Routers only advertise their existence, not the best way to reach a given destination network. If a host uses a non-optimal route, ICMP Redirect messages redirect the host to the better route.

Table 3.7 describes the settings for ICMP Router Discovery.

Table 3.7 ICMP Router Discovery Settings

Router Discovery Setting	Description
Level of Preference	The preference level of this router to be the default gateway. The default value is 0.
Advertisement Lifetime (minutes)	The time, in minutes, after which a host considers a router down (after it has received its last Router Advertisement). The default time is 30 minutes.
Advertisement interval minimum time (minutes)	The minimum amount of time between Router Advertisements sent by this router. The default is 7 minutes.
Advertisement interval maximum (minutes)	The maximum amount of time between Router Advertisements sent by this router. The default is 10 minutes. Router Advertisements are sent at a random interval between the minimum and maximum times.

To enable router discovery advertisements

1. In the **Routing and Remote Access** snap-in, open the **IP Routing** container for the desired server.
2. Open the **General** container, right-click the desired interface, and click **Properties**.
3. From the **General** tab, select the **Enable router discovery advertisements** check box.

Windows 2000, with the **Enable router discovery advertisements** option selected, sends Router Advertisements periodically and in response to Router Solicitations using the IP multicast address of 224.0.0.1. TCP/IP for Windows 2000 and Windows 98 supports the use of Router Solicitation messages to discover the default gateway. For information about how to disable router discovery for TCP/IP for Windows 2000, see "Windows 2000 TCP/IP" in the *TCP/IP Core Networking Guide*.

Additional Resources

For more information about IP routing, see:

- *Routing in the Internet* by Christian Huitema, 1995, Englewood Cliffs, NJ: Prentice Hall.
- *OSPF: Anatomy of an Internet Routing Protocol* by John T. Moy, 1998, Reading, MA: Addison-Wesley.
