UnixWorld Online: Technical Feature Article: No. 002

To RIP or To OSPF?

That is the question: should you use the older, more established RIP (Routing Information Protocol) or the new-kid-on-the-block OSPF (Open Shortest Path First) IGP (Interior Gateway Protocol)? This article can help you make the right decision--it describes the pro's and con's for using either protocol and provides some specific recommendations.

By Eddie Rabinovitch.

The ''Subnet Addressing'' tutorial by Ron Cooney gives a detailed overview of the IP subnetworking scheme. This article demystifies IP subnetworking further and provides some insights on how dynamic routing is done in the IP world, by describing the more popular routing techniques.

The routing algorithm in an IP network is straightforward: if the network and subnetwork part match a directly connected network, then send the IP datagram to the destination IP address on that subnet (resolving the complete IP address into a physical address). Otherwise, examine the routing table to find the longest match to the destination IP address. If a matching entry is found, route to the indicated IP address (next hop). If no match found then discard the packet (and send Destination Unreachable message).

TCP/IP products can use static route definitions or a dynamic routing protocol, such as RIP (Routing Information Protocol), OSPF (Open Shortest Path First), Border Gateway Protocols, and other dynamic routing schemes that exchange routing information to resolve routing errors. This article compares the two most important routing mechanisms for interior (intra-domain) routing: RIP and OSPF. It does not include a comparison with the more advanced (but less popular) modification of RIP, RIP II, described in RFC 1387 (6K text file), RFC 1388 (16K text file), and RFC 1389 (24K text file).

RIP

RIP is an Interior Gateway Protocol (originally defined by RFC 1058 (90K text file) is the most widely accepted routing protocol. It is also known by the name of the Unix daemon program `routed`, which was originally designed at U.C. Berkeley to provide consistent routing and reachability information among machines on a local area network. RIP's popularity is not necessarily based on its technical merits, but probably results because U.C. Berkeley distributed `routed` along with their popular 4.x BSD UNIX systems. Thus, many Internet sites adopted and installed `routed` and started using RIP without even considering its technical merits and limitations. Once installed and running, it became the basis for local routing.

RIP is straightforward: it arranges to have routers to broadcast their entire current routing database periodically, typically every 30 seconds. This message lists each destination along with a ''distance'' to

that destination measured in number of router hops. RIP is also known as a distance-vector routing algorithm, which means that there is a distance (a cost) and a vector (a direction) for each destination (the vector just shows the name of the neighboring router, not the entire path).

As a distance-vector based algorithm, RIP works fine for small, stable high-speed networks. Instead of passing along the status of the links to the networks, the router tells its neighbors about the entire world, where the ‘‘best link’’ is precomputed, not necessarily the fastest one. Here, ‘‘best’’ means a route with the least number of gateway hops, relative to the active routing table.

In times of network instability--because every router is broadcasting his entire routing table--it takes awhile for those states to converge into a common view of the network topology. Specifically, RIP does not address well these major types of problems:

- There is no protection from routing loops within RIP based networks. Therefore, the implementation must trust all network participants to prevent such loops.
- RIP uses a hop count of 15 do denote infinity, which makes it unsuitable for large networks.
- RIP has the so called slow convergency or count-to-infinity problem, in which inconsistencies arise because routing update message propagate slowly across the network. Particularly, in large networks (or networks with slow links), some routers may still advertise a route that has vanished. (That, by the way, was one of the reasons 15 was chosen as the value of infinity to limit the slow convergency).

Slow convergency can be addressed by a technique called hold down, which forces the router to ignore information about a network for a fixed period of time (typically, 60 seconds). The idea is to wait long enough so that all machines receive the bad news of the vanished link and do not mistakenly accept outdated information. The disadvantage of this technique is that incorrect routes and routing loops will be preserved for the duration of the hold down, even when a valid alternate path is available.

Another, more popular approach to address the slow convergency problem is a technique known as split-horizon update. This technique is widely used by router vendors. With split horizon, a router records the interface over which it received a particular route and does not propagate its higher-cost route back over the same interface. However, split horizon does not resolve the slow convergency problem for all topologies. It also introduces a problem for a frame relay network that would not permit full-meshed connectivity for partially meshed networks.

RIP allows the use of subnet masks with static information defined in the routers. RIP has no provision to exchange subnet mask information between routers. The routers in your network should use the same subnet masks and RIP to make IP datagrams travel from source to destination within your network. This is another technical limitation of RIP. For example, with frame relay, the entire frame relay network can be assigned one subnet, but as was mentioned above, because of the split horizon technique it would not allow connectivity between two remote sites via a centrally based router, and require explicit definitions of additional routes for all remote sites.

If the subnet mask is not constant in a RIP based network, it becomes impossible for the routers to distinguish between network part and host part because RIP cannot dynamically update or change the subnet mask. Hence one mask should be used throughout the subnetwork. If a variable subnet mask is needed, static routes or more advanced routing protocols, for instance, OSPF, have to be used.

OSPF

OSPF, as defined by RFC 1131 (781K PostScript file), is a link-state algorithm. In contrast to a distance-vector algorithm, where a router ‘‘tells all neighbors about the world,’’ link-state routers ‘‘tell the world about the neighbors.’’ OSPF specifies a class of messages called link-state advertisements (LSAs) that allow routers to update each other about the LAN and WAN links to which they are connected. When a change is made to the network, LSAs flow between routers.

OSPF routers receive link-state updates and store them in a topology database in memory. The typical OSPF database contains a representation of every link and router in the enterprise network. When routers receive internetwork traffic that needs to be forwarded towards a destination end node, they use their topology database to calculate a table of the best routes through an Internet.

OSPF addresses all RIP shortcomings and thus is better suited for modern large, dynamic networks. For example, in contrast to RIP sending the entire routing table from router to router every 30 seconds, OSPF sends its link state information every 30 minutes. OSPF can get away with this, because OSPF routers also send each other small update messages (typically less than 75 bytes), whenever they detect a change in the network (for instance, a failure or a new link). When routers exchange updates that reflect changes in the network, they ‘‘"converge’’ on a new representation of the topology quickly and accurately.

Although, it addresses all of the problems with RIP, OSPF itself is not an absolutely perfect routing protocol (if there is such a thing?). For small and medium-size networks, the basic services of OSPF works very well, enabling a wide range of robust TCP/IP and multiprotocol topologies. But in really large configurations, the huge number of router updates--that flow between routers--can become an issue. For instance, in a mesh network with over a hundred routers a single link change can precipitate a flood of thousands of link-state messages that propagate across the entire network from router to router. In each router, the database that stores these messages can grow to over a megabyte of live data. Each time there is a change to the network, routers must recalculate new routes. In very large OSPF networks, topology convergence can be delayed, while routers exchange link-state messages, update databases, and recalculate routes.

This delay in network convergence is the natural effect of very large topologies, and it will occur with any router protocol. Fortunately, OSPF was designed and implemented in a much better way than RIP to address this issue by what’s know as the OSPF area. OSPF areas are simply logical subdivisions of an OSPF network. Typically, an enterprise is divided into areas that correspond to buildings, campuses, regions, or other administrative domains. An enterprise can have a practically unlimited number of areas.

OSPF routers within one area do not exchange topology updates with the routers in the other areas. When a LAN or WAN link is added to one area, topology updates flow only to routers within that area. This reduces the number of updates that flow through the network and the size of the topology databases in each router. In an enterprise with 500 routers, the creation of 10 areas of 50 routers means that each router only needs to store link information for 50 routers, not 500.

OSPF areas are connected to each other by means of a backbone that is just another area unto itself. A router that connects its area to the backbone must maintain a topology database for both areas. These special "multi-area" routers are called area border routers, and they serve as a filter for topology updates

that move between areas and the backbone. Area border routers communicate with each other using special link-state messages that contain a short-hand summarization of the LAN or WAN topology in their areas.

Area border routers summarize topology information with an addressing technique that is analogous to the U.S. Post Office ZIP codes. When a post office in New York City, for instance, handles a letter addressed to San Francisco, it only needs to look at the first few digits in the ZIP code to know that the letter is bound for San Francisco; the remaining digits are not relevant at that point in the process. The U.S. phone-switching system works the same way: when routing a call, major phone switches decode just the three-digit area-code prefix of a phone number to determine which long-distance trunk line to select; the rest of the digits are ignored until the call is routed to the right state and city.

These techniques of ‘‘hierarchical’’ addressing reduce the complexity of the phone and postal systems, because routing elements do not need to know all the details of end-to-end routes to every possible destination. Route information at the top of the hierarchy only relates to regions or areas. The OSPF area feature works the same way. Area border routers send-out special LSA update messages that advertise a range of IP addresses that reside in an area. Area border routers store these summarization messages in a special database that tells them how to forward inter-area traffic between areas. Route calculations based on summarization only need to determine what range an IP destination address falls within, based on the first few bytes of the address, not the complete address. However, this addressing feature does not come for free: before this process begins, OSPF summarization parameters must be administratively configured in area border routers. This task is similar to the configuration of router traffic filters or priorities.

A good example of OSPF areas’ benefits can be seen in a campus environment, where each building is defined as an area. For example, lets take a campus where each building has 12 floors and a multiprotocol router on each floor. Without OSPF areas, routers would have to exchange updates with every other router on the campus, creating, in the process, topology databases that represent every routing node and link. When areas are deployed, routers only exchange link state information with routers in the same building. An area border router in each building forms a link between the building and the campus backbone.

Another OSPF area scenario is a national internetwork that is divided into areas that correspond to different regions of the country. For example, all the routers in the New York area would have identical databases that cover the New York region only, and the same would apply to other regions as well, Chicago, Tampa, and so forth. In each area, an area border router is attached to the national backbone. This approach eliminates the need to propagate router update messages across entire national (or international) Internets.

The area feature of OSPF is not the only factor in building large reliable networks, but it is one of the most important. When deploying OSPF areas, it is critical that area border routers have enough resources (memory, processing power) to handle topology chores for both the local and the backbone areas. Low-end routers may be overwhelmed by the demands of maintaining multiple databases.

Some additional considerations to keep in mind:

- Choose the correct topology (hub, mesh, hierarchical, collapsed backbone, and so forth).
- Fine-tune OSPF timers within routers. The frequency of OSPF administrative messages often can

be adjusted, and the right settings conserve bandwidth and speed convergence.

- Use variable subnet addressing to maximize the number of viable IP addresses within an organization; in TCP/IP networks, the 32 bits of an IP address are divided up into parts that identify an end-node and its attached network; variable subnet addressing allows the 32 bits to be divided differently for different areas of a network, which allows more nodes, networks, and subnetworks to be addressed, given a finite address space.
- Assign a meaningful OSPF ‘‘cost’’ to each link. OSPF computes its routes based on a least-cost metric: in configuring a network, each link is assigned a relative cost, based on its bandwidth, security, reliability, and so on. The ability of an OSPF network to adapt to end-user demands is directly proportional to the care invested in assigning cost values to links.

Summary

In many places, RIP is still used in TCP/IP networks that have not been upgraded to OSPF. It is also used on OSPF networks as an end-station-to-router protocol. OSPF addresses all the deficiencies of RIP, without affecting connectivity to RIP based networks. Fast growing networks must be designed properly if the capabilities of OSPF are to be fully exploited. Because of its ability to handle variable networking masks, OSPF also helps to reduce waste of today’s precious IP addresses. Ideally, network design should include a consistent enterprise-wide IP address assignment policy that lends itself to the creation of OSPF areas and address summarization. If correct design and router-tuning takes place, OSPF will allow networks to scale to very large topologies, while maintaining high levels of availability and performance.

Copyright © 1995 The McGraw-Hill Companies, Inc. All Rights Reserved.
Edited by Becca Thomas / Online Editor / UnixWorld Online / beccat@wcmh.com

Go To Content

Search

Last Modified: Saturday, 04-Nov-95 07:07:15 PST