APPENDIX G

Which Routing Protocol?

Among all the thorny questions that network engineers are asked on a regular basis, probably among the hardest is this one:

> My network currently runs Enhanced Interior Gateway Routing Protocol (EIGRP). Would I be better off if I switched to Open Shortest Path First (OSPF)?

You can replace the two protocols mentioned in this sentence with any pair of protocols among the advanced interior gateway protocols (OSPF, Intermediate System-to-Intermediate System [IS-IS] and EIGRP), and you have described a question that routing protocol engineers are asked probably thousands of times a year. Of course, convergence is always faster on the other side of the autonomous system boundary, so to speak, so it is always tempting to jump to another protocol as soon as a problem crops up with the one you are running.

How do you answer this question in real life? You could try the standard, "It depends," but does this really answer the question? The tactic in the Routing Protocols Escalation Team was to ask them questions until they went away, but none of these answers really helps the network operator or designer really answer the question, "How do you decide which protocol is the best?"

Three questions are embedded within this question, really, and it is easier to think about them independently:

- Is one protocol, in absolute terms, "better" than all the other protocols, in all situations?
- If the answer to this first question is "No," does each routing protocol exhibit some set of characteristics that indicate it would fit some situations (specifically, network topologies) better than others?
- After you have laid out the basics, what is the tradeoff in living with what you currently have versus switching to another routing protocol? What factors do you need to consider when doing the cost/benefit analysis involved in switching from one routing protocol to another?

This appendix takes you through each of these three questions. This might be the first and last time that you hear a network engineer actually answer the question, "Which routing protocol should I use?" so get ready for a whirlwind tour through the world of routing.

Is One Protocol "Better" Than the Others?

The first thing you need to do with this sort of question is to qualify it: "What do you mean by better?" Some protocols are easier to configure and manage, others are easier to troubleshoot, some are more flexible, and so on. Which one are you going to look at?

This appendix examines ease of troubleshooting and convergence time. You could choose any number of other measures, including these:

- **Ease of management**—What do the Management Information Bases (MIBs) of the protocol cover? What sorts of **show** commands are available for taking a network baseline?
- **Ease of configuration**—How many commands will the average configuration require in your network configuration? Is it possible to configure several routers in your network with the same configuration?
- **On-the-wire efficiency**—How much bandwidth does the routing protocol take up while in steady state, and how much could it take up, at most, when converging in response to a major network event?

Ease of Troubleshooting

The average uptime (or reliability) of a network is affected by two elements:

- How often does the network fail?
- How long does it take to recover from a failure?

The network design and your choice of equipment (not just the vendor and operating system, but also putting the right piece of equipment into each role and making certain that each device has enough memory, and so on) play heavily into the first element. The design of the network also plays into the second element. The piece often forgotten about when considering the reliability of a network is how long it takes to find and fix, or troubleshoot, the network when it fails.

Ease of management plays a role in the ease of troubleshooting, of course; if it is hard to take a baseline of what the network is supposed to look like, you will not do so on a regular basis, and you will have a dated picture to troubleshoot from. The tools available for troubleshooting are also important. Of course, this is going to vary between the implementations of the protocols; here, implementations in Cisco IOS Software illustrate the concepts. Table G-1 outlines some of the troubleshooting tools that are available in EIGRP, OSPF, and IS-IS, in Cisco IOS Software.

Table G-1 *Cisco IOS Software Troubleshooting Tools for EIGRP, OSPF, and IS-IS*

	EIGRP	OSPF	IS-IS
Debug Neighbors	Neighbor formation state; hello packets.	Neighbor formation state; hello packets.	Packets exchanged during neighbor formation.
Log Neighbor State	Yes.	Yes.	No.
Debug Database Exchange and Packets	Packets exchanged (updates, replies, and so on), with filters per neighbor or for a specific route.	Packets flooded, with filters for specific routing information. Packets retransmitted.	Packets flooded.
Debug Interactions with the Routing Table	Yes.	No.	No.
Debug Route Selection Process	Yes (DUAL[1] FSM[2] events).	Yes (SPF[3] events).	Yes (SPF events).
Show Database	Yes, by specific route and route state.	Yes, by LSA[4] type and advertising router.	Yes, by LSP[5] ID or type of route.
Event Log	Yes; understandable if you comprehend DUAL and its associated terminology.	Yes; only understandable if you have access to the source code.	No.

[1] DUAL = Diffusing Update Algorithm

[2] FSM = finite state machine

[3] SPF = shortest path first

[4] LSA = link-state advertisement

[5] LSP = link-state packet

From this chart, you can see that EIGRP generally provides the most tools for finding a problem in the network quickly, with OSPF running a close second.

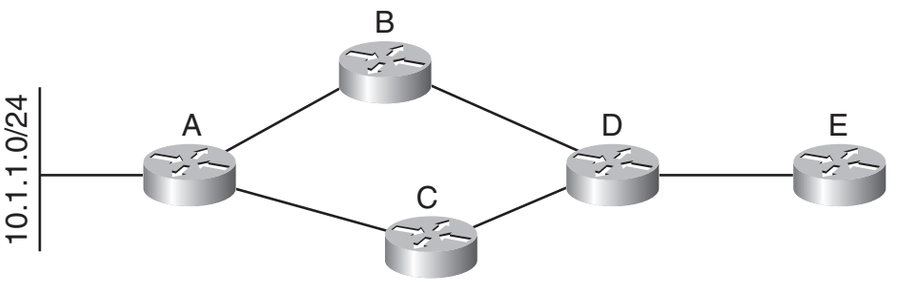
Which Protocol Converges Faster?

I was once challenged with the statement, "There is no way that a distance vector protocol can ever converge faster than a link-state protocol!" An hour and a half later, I think the conversation tapered off into, "Well, in some situations, I *suppose* a distance vector protocol *could* converge as fast as a link-state protocol," said without a lot of conviction.

In fact, just about every network engineer can point to reasons why he thinks a specific routing protocol will *always* converge faster than some other protocol, but the reality is that all routing protocols can converge quickly or slowly, depending on a lot of factors strictly related to network design, without even considering the hardware, types of links, and other random factors that play into convergence speed in different ways with each protocol. As a specific

example, look at the small network illustrated in Figure G-1 and consider the various options and factors that might play into convergence speed in this network.

Figure G-1 *Simple Network*



This figure purposefully has no labels showing anything concerning routing protocols configuration or design; instead, this section covers several possible routing configurations and examines how the same protocol could converge more or less quickly even on a network this small through just minor configuration changes.

Start with EIGRP as an example:

- The Router A to C link has a bandwidth of 64 kbps.
- The Router A to B link has a bandwidth of 10 Mbps.
- The Router B to D and Router C to D links have equal bandwidths.

With this information in hand, you can determine that Router D is going to mark the path to 10.1.1.0/24 through Router B as the best path (the *successor* in EIGRP terms). The path through Router C will not be marked as a *feasible successor*, because the differential in the metrics is too great between the two paths. To the EIGRP process running on Router D, the path through Router C cannot be proven based on the metrics advertised by Routers B and C, so the path through Router C will not be installed as a possible backup route.

This means that if the Router B to D link fails, Router D is forced to mark 10.1.1.0/24 as *active* and send a query to Router C. The convergence time is bounded by the amount of time it takes for the following tasks:

- Router D to examine its local topology table and determine that no other known loop-free paths exist.
- Router D to build and transmit a query toward Router C.
- Router C to receive and process the query, including examining its local EIGRP topology table, and find it still has an alternate path.
- Router C to build a reply to the query and transmit it.
- Router D to receive the reply and process it, including route installation time and the time required to change the information in the forwarding tables on the router.

Many factors are contained in these steps; any one of them could take a long time. In the real world, the total time to complete the steps in this network is less than two or three seconds.

Now change the assumptions just slightly and see what the impact is:

- The Router A to C link and A to B links have equal bandwidth.
- The Router B to D link has a bandwidth of 64 kbps.
- The Router B to C link has a bandwidth of 10 Mbps.

As you can tell, the network conditions have been changed only slightly, but the results are altered dramatically. In this case, the path to 10.1.1.0/24 through Router C is chosen as the best path. EIGRP then examines the path through Router B and finds that it is a loop-free path, based on the information embedded in EIGRP metrics. What happens if the Router B to C link fails?

The process has exactly one step: Router D examines its local EIGRP topology table and finds that an alternate loop-free path is available. Router D installs this alternate route in the local routing table and alters the forwarding information as needed. This processing takes on the order of 150 milliseconds or less.

Using the same network, examine the various reactions of OSPF to link failures. Begin with these:

- The Router B to D link has a cost of 20.
- All other links in the network have a cost of 10.
- All routes are internal OSPF routes.

What happens if the Router B to C link fails?

1. Router B and C detect the link failure and wait some period of time, called the link-state advertisement (LSA) generation time. Then they flood modified router LSAs with this information.
2. The remaining routers in the network receive this new LSA and place it in their local link-state databases. The routers wait some period of time, called the shortest path first (SPF) wait time, and then run SPF.
3. In the process of running SPF, or after SPF has finished running (depending on the implementation), OSPF will install new routing information in the routing table.

With the default timers, it could take up to one second (or longer, in some situations) to detect the link failure and then about three and a half seconds to flood the new information. Finally, it could take up to two and a half seconds before the receiving routers will run SPF and install the new routing information. With faster times and various sorts of tuning, you can decrease these numbers to about one second or even in the 300-millisecond range in some specific deployments.

Making Router D an area border router (ABR) dramatically impacts the convergence time from the Router E perspective because Router D has to perform all the preceding steps to start

convergence. After Router D has calculated the new correct routing information, it must generate and flood a new summary LSA to Router E, and Router E has to recalculate SPF and install new routes.

Redistributing 10.1.1.0/24 into the network and making the area that contains Routers A, B, C, and D into a not-so-stubby area (NSSA) throws another set of timers into the problem. Router D now has to translate the Type 7 external LSA into an external Type 5 LSA before it can flood the new routing information to Router E.

These conditions do not even include the impact of multiple routes on the convergence process. EIGRP, for instance, can switch from its best path to a known loop-free path for 10,000 routes just about as fast as it can switch 1 route under similar conditions. OSPF performance is adversely impacted by the addition of 10,000 routes into the network, possibly doubling convergence time.

You can see, then, that it is not so simple to say, "EIGRP will always converge faster than OSPF," "IS-IS will always converge faster than EIGRP," or any other combination you can find. Some people say that OSPF always converges faster than EIGRP, for instance, but they are generally considering only intrarea convergence and not the impact of interarea operations, the impact of various timers, the complexity of the SPF tree, and other factors. Some people say that EIGRP always converges faster than any link-state protocol, but that depends on the number of routers involved in the convergence event. The shorter the query path, the faster the network converges.

If you align all the protocol convergence times based on the preceding examination, you generally find the convergence times in this order, from shortest to longest:

1 EIGRP with feasible successors.

2 Intrarea OSPF or IS-IS with fast or tuned timers.

3a EIGRP without feasible successors.

3b Intrarea OSPF or IS-IS with standard timers.

3c Interarea OSPF or IS-IS.

The last three are highly variable, in reality. In any particular network, OSPF, IS-IS, and EIGRP without feasible successors might swap positions on the list. The network design, configuration, and a multitude of other factors impact the convergence time more than the routing protocol does. You get the best convergence time out of a routing protocol if you play the network design to the strengths of the protocol.

Which Designs Play to the Strength of Each Protocol?

The natural question, after you have decided that network design plays into the suitability of the protocol (you have seen this to be the case for convergence speed, but the same is also true of

any other factor you might consider for a given routing protocol, including management, troubleshooting, configuration, and so on) is this:

What sorts of network designs play into the strengths of any given routing protocol?

This is not an easy question to answer because of the numerous ways to design a network that works. Two- and three-layer network designs, switched cores versus routed cores, switched user access versus routed user access—the design possibilities appear to be endless. To try to put a rope around this problem, the sections that follow examine only a few common topological elements to illustrate how to analyze a specific topology and design and try to determine how a routing protocol will react when running on it.

The specific types of network topologies considered here are as follows:

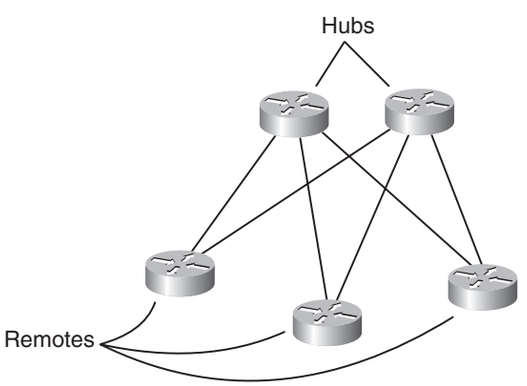
- Hub-and-spoke designs
- Full mesh designs
- Highly redundant designs

After you consider each of these specific topology elements, you learn the general concepts of hierarchical network design and how each protocol plays against them.

Hub-and-Spoke Topologies

Hub-and-spoke network designs tend to be simple in theory and much harder in implementation. Scaling tends to be the big problem for hub-and-spoke topologies. The primary focus here is the capability of a routing protocol to maintain a multitude of routing neighbors and to converge to massive network events in an acceptable amount of time. Assume, throughout this section, that you are always dealing with dual-homed hub-and-spoke networks, as Figure G-2 illustrates.

Figure G-2 *Dual-Homed Hub-and-Spoke Network*



Start by considering the following simple question:

> How many spokes or remote routers does it take to really start stressing any routing protocol that is running over a hub-and-spoke network design?

The answer to this question always depends on various factors, including link speed and stability, router processing speed and packet switching speeds, and other factors. However, general experience shows that a high-speed router (in terms of processing power) with reasonably good design supports at least 100 remote sites with any modern routing protocol.

When considering network designs in which hundreds of remote sites are available, however, you need to use special techniques with each protocol to scale the number of remote sites attached to a single pair of hub routers. Look at each protocol to see what types of problems you might encounter and what types of tools are available to resolve those problems:

- OSPF floods topology information to each router within an area and summaries of reachability information into the area. You can place all the remote site routers into one or more OSPF *stub areas*, which cuts down on the amount of information flooded out to each remote site. Any change on a remote site is still flooded to every other remote site within the same area. For that reason, the design becomes a tradeoff between the number of areas that you want to manage and that the hub routers support and the amount of information that you can flood through the low-speed links connecting the remote stub sites.
- IS-IS also floods information to each router within an area. It does not, by default, flood information from the core of the network (the L2 routing domain) into each area. Again, you still face the tradeoff of how many level 1 routing domains you want to support at the hub routers versus how much information you can flood toward each remote router.
- The primary factor in determining scaling and convergence time in an EIGRP hub-and-spoke network is the number of queries the hub router needs to generate or process when the network changes, and the number of updates the hub router needs to generate toward the remote. Normally, if a hub loses several routes, for instance, it needs to generate queries for each of those routes to each of the remote sites. The remote sites then query the other hub router, which must process and reply to each of the queries. If the number of routes is high, this can be a processor- and memory-intensive task, causing the network to converge slowly, especially if the links between the remote sites and the hub routers are low speed. In this situation, you can summarize routers at the core toward the remote routers and block the routing information transmitted up toward the core routers. You can also cut down on the query range into the hub-and-spoke network dramatically. EIGRP, however, also provides a special operational mode for the remote sites; you can configure the remote sites as *stubs*, which indicates to the hub routers that the remote sites are never used for transiting traffic. If the remote sites are configured as stub routers, the hub router never queries them for lost routes, and the scaling properties change dramatically.

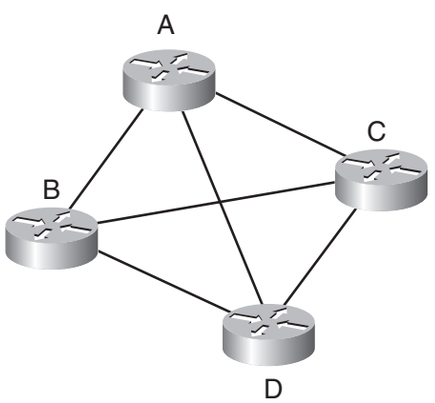
EIGRP, in theory, scales much better in a hub-and-spoke topology—and this is true in real networks, too. You often find EIGRP hub-and-spoke networks that have more than 500 remote sites attached to a pair of hub routers, over low bandwidth links, in the wild. In contrast, you

tend to see OSPF and IS-IS hub-and-spoke networks top out at around 200 remote sites, even if higher bandwidth links are involved.

Full Mesh Topologies

Full mesh topologies are a less common design element in networks, but they are worth considering because the scaling properties of a routing protocol in a full mesh design indicate, to some degree, the scaling properties of the same protocol in a partial mesh design. You can think of a full mesh topology as a special case of a partial mesh topology. Again, look at the challenges and tools that are available for each protocol. Use the network illustrated in Figure G-3 throughout this discussion.

Figure G-3 *Full Mesh Network*



- Each OSPF router sends topology information to each adjacent neighbor within an area (flooding domain). If Router A receives a new link-state advertisement (LSA), Router D receives three copies of this new LSA: one from Router A, one from Router B, and one from Router C. The Cisco IOS Software implementation of OSPF does have an option to control the flooding through a full mesh network, using the **database filter-out** command.
- IS-IS is similar to OSPF; each router sends topology information to each adjacent neighbor. Cisco IOS Software enables you to control flooding through *mesh groups*.
- Each router in an EIGRP network sends each of the routes it is using to forward traffic to each neighbor. In this network, Router D is going to receive three copies of any new routing information that Router A receives, one copy from Router A, one from Router B, and one from Router C. These three copies of the routing information might be the same, but they indicate reachability through three different next hops (or neighbors). Reducing the information propagated through the mesh is difficult, at best. You can filter these routing updates through some paths within the mesh to decrease the amount of information flooded through the mesh, but that also reduces the number of paths usable through the mesh for any specific destination.

OSPF and IS-IS flood extra information through a mesh topology by default, but you can use tools to reduce the amount of flooding in highly meshed topologies. EIGRP sends updates through each router in the mesh, but it is difficult to reduce the number of these updates unless you want to decrease the number of paths that the network actually uses through the mesh.

In the real world, OSPF and IS-IS scale better in highly meshed environments, especially if you implement flooding reduction techniques. This is a matter of scale, of course; networks that have a mesh network of 20 or 30 routers work fine with any of the three routing protocols. However, when the mesh starts surpassing this number of routers, the special techniques that OSPF and IS-IS offer to scale further can make a difference.

Interaction with Hierarchical Designs

Traditional network design is based on layers, either two or three, that abstract the network details into "black boxes" and divide functionality vertically through the network to make management and design easier:

- The two-layer model has *aggregation* and *core layers*, or *areas*, within the network.
- The three-layer model has *access*, *distribution*, and *core layers*.

How do these layered network designs interact with each protocol? Consider each protocol in turn.

OSPF splits flooding domains into areas that are separated by ABRs. Because every router within an area must share the same link-state database to calculate loop-free paths through the network, the only place that route aggregation can be performed is at an ABR. ABRs actually aggregate two types of information:

- Information about the topology of an area that is hidden from other areas at these border edges
- Aggregation of reachability information that can be configured at these border edges

This combination of route aggregation points and flooding domain boundaries in the network implies several things:

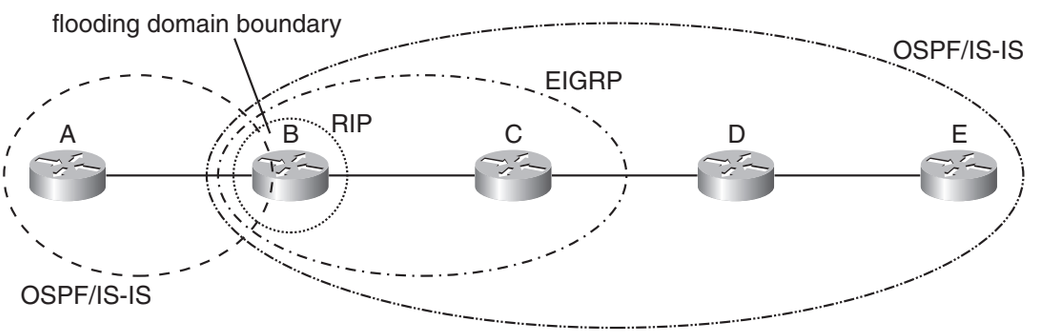
- In all three-layer network designs with OSPF, you should place the ABR in the distribution layer of the network.
- In all two-layer network designs with OSPF, you should place the ABR at the aggregation to core layer edge of the network.
- The most aggregation points that you can cross when passing from one edge of the network to the opposite edge of the network is two.

These topological limitations might not be major in smaller networks, but in networks that have thousands of routers, they could impose severe restrictions on the network design. Network designers and operators normally break up OSPF networks at this size into multiple administrative domains, connecting the separate domains through BGP or some other mechanism.

IS-IS is similar to OSPF in its restrictions, except that IS-IS allows the core and outlying flooding domains to overlap. This introduces a degree of flexibility that OSPF does not provide, but you can still only aggregate routing information at the edges where two flooding domain meet, and you cannot build more than two levels of routing into the network.

EIGRP, as a distance vector protocol, does not divide the concepts of topology summarization and routing aggregation; topology beyond one hop away is hidden by the natural operation of the protocol. Figure G-4 illustrates the conceptual difference among EIGRP, OSPF/IS-IS, and RIP in terms of topology information propagated through the network.

Figure G-4 *Topological Awareness in Routing Protocols*



If you examine the scope through which routing information is transmitted (or known) within a network, you find the following:

- The Bellman-Ford algorithm, used by the Routing Information Protocol (RIP) and the Interior Gateway Routing Protocol (IGRP), uses only information about the local cost to reach a given destination. If Router B is running RIP, it considers only the total cost of the path to reach a destination at Router E when deciding on the best (loop-free) path.
- Diffusing Update Algorithm (DUAL), used by EIGRP, considers the local cost to reach a given destination and the cost of each neighbor to reach the same destination when calculating which available paths are loop free. EIGRP uses an awareness of the topology that is one hop away from the calculating router.
- OSPF and IS-IS, which are link-state protocols, do not use information about the metrics of a neighbor; rather, they count on being aware of the entire topology when calculating a loop-free path. At a flooding domain border, OSPF and IS-IS act much like distance vector protocols. Router A does not know about the topology behind Router B; it only knows the cost of Router B to reach destinations that are attached to Router E.

Because topology information is hidden in the natural processing of EIGRP routing updates, EIGRP is not restricted in where it can aggregate routing information within the network. This provides a great deal of flexibility to network designers who are running EIGRP. Multiple layers of aggregation can be configured in the network. This means that moving from one edge of the

network to the opposite edge of the network could mean encountering many more than two aggregation points.

The practical result of the EIGRP capability to aggregate routing information anywhere in the network is that many existing large-scale (2000 router and larger) networks run within a single EIGRP process or administrative domain. The feasibility of building networks this large is based on the capability to use route aggregation to divide the network into multiple layers, or sections, each acting fairly independently of the other. Although it is possible to build an OSPF or IS-IS network this large, designing and managing this network is more difficult because of the restrictions that link-state protocols place on aggregation points.

In general, up to some relative size, the protocols are relatively equal in their capability to work with hierarchical network designs. OSPF and IS-IS tend to be less flexible about where route aggregation can be placed in the network, making it more difficult, in some situations, to fit the network design and the protocol design together. EIGRP excels at fitting into hierarchical network design.

Topological Rules of Thumb

After examining these various network topologies and how each routing protocol tends to react, you can see that when a network does not reach the edge of a specific protocol capability on any given topology, any of the routing protocols is fine. If your network has a specific predominant topology type, however, such as large-scale hub-and-spoke or large-scale full mesh topologies, choosing a protocol to fit those topologies makes sense. You can always compromise in complex areas of your network design by making effective and stable topological design areas in which the routing protocol is really stretched to the edge of its capabilities.

What Are the Tradeoffs?

In many networks, the final decision of which routing protocol is "best" comes down to these issues:

- **Convergence speed**—How important is convergence speed? How much flexibility do you have in the design of your network around convergence speeds?
- **Predominant topologies**—Does your network design have one dominant type of topology? Would a full mesh or large-scale hub-and-spoke topology benefit from running one protocol over another?
- **Scaling strategy**—Does your scaling strategy call for dividing the network into multiple pieces, or does it call for a single IGP domain, with the network broken up into pieces through route aggregation and other techniques?
- **Maintenance and management**—Which routing protocol fits the network management style of your day-to-day operations? Which one seems easier to troubleshoot and manage in your environment?

Beyond the technical factors are some nontechnical ones. For instance, if you decide to switch protocols, what is the cost for the long term? You need to consider training costs, the cost of revised procedures, design effort, and possible downtime while you convert the network from one protocol to another.

In some situations, this might not be an issue. For instance, if two networks are being merged because of a corporate merger, and each runs a different protocol, the decision might be more open to consideration. If you are going to need to convert one half of the network or the other, you can more carefully consider the technical considerations and make a decision based on those considerations alone. However, if your network is stable today, you should think twice about switching protocols unless a change in the business environment or some major shift in the way the network is built indicates it is an important move to make to meet the needs of the enterprise.