Enterprise SOLUTIONS

ILLUSTRATION BY JOHN RITTER

Conduits for Content

A Guide to Planning and Designing Enterprise Content Delivery Networks

The ability of enterprises to deliver rich media content—streaming video, audio, animation, and heavy graphics—across a national or global private intranet or through a service provider content-hosting service enables new applications and richer company communications. It also optimizes the delivery of more traditional static content such as HTML code and text in Web pages. With a new generation of products and technologies, Cisco has simplified the task of designing end-to-end content delivery networks.

The challenge of delivering content today is similar to the challenges networks faced in the early days of enterprise routing. "We're applying the same principles to content delivery as we applied years ago to scaling the Internet," says Joe Hielscher, Marketing Director in

How Content Delivery Networks Add Value

The Cisco content delivery network provides these key benefits:

- Improved access performance, because content is placed topologically closer to users. This element is especially important for networks whose end users increasingly deploy rich media services, such as streaming video, on their Web sites.
- Higher content availability, because content is replicated and distributed across the network in multiple locations, thereby avoiding single-site or single-server failures.
- Improved management control over content replication and distribution, resulting in applications being optimized to deliver the appropriate service levels for preselected locations in a value-added network.

Cisco's Content Services Business Unit. "We're looking at principles of hierarchy, distribution, and intelligence in the network. But for the first time, we're applying the ideas we used to build a Layer 3 Internet interconnectivity fabric to build a Layer 7-aware Internet; *that's* what content delivery networks are about."

Cisco Content Delivery Networks

Through a combination of internally developed technology and the acquisitions of ArrowPoint Communications, Inc. and SightPath, Inc.—two leading developers of content delivery technology—Cisco products now enable both enterprises and service providers to build end-to-end content delivery networks.

Content delivery networks are value-added networks built on top of existing intranet or Internet infrastructures where content is replicated on intelligent caches and placed topologically closer to the users requesting it. A Cisco content delivery network combines traditional routing and switching intelligence with content-aware technology at both enterprise data centers or service provider distribution points. The network determines the best site to serve the content and, within the site, the best server or content engine for the requested content, by actively probing or passively monitoring the changing characteristics of the network. The network then directs the requests based on a combination of metrics. By localizing the repeated content transmission through the intranet—a common performance bottleneck on networks—a content delivery network lowers propagation delays and latency and improves overall response time. These results translate into higher levels of user satisfaction and productivity.

Design Considerations

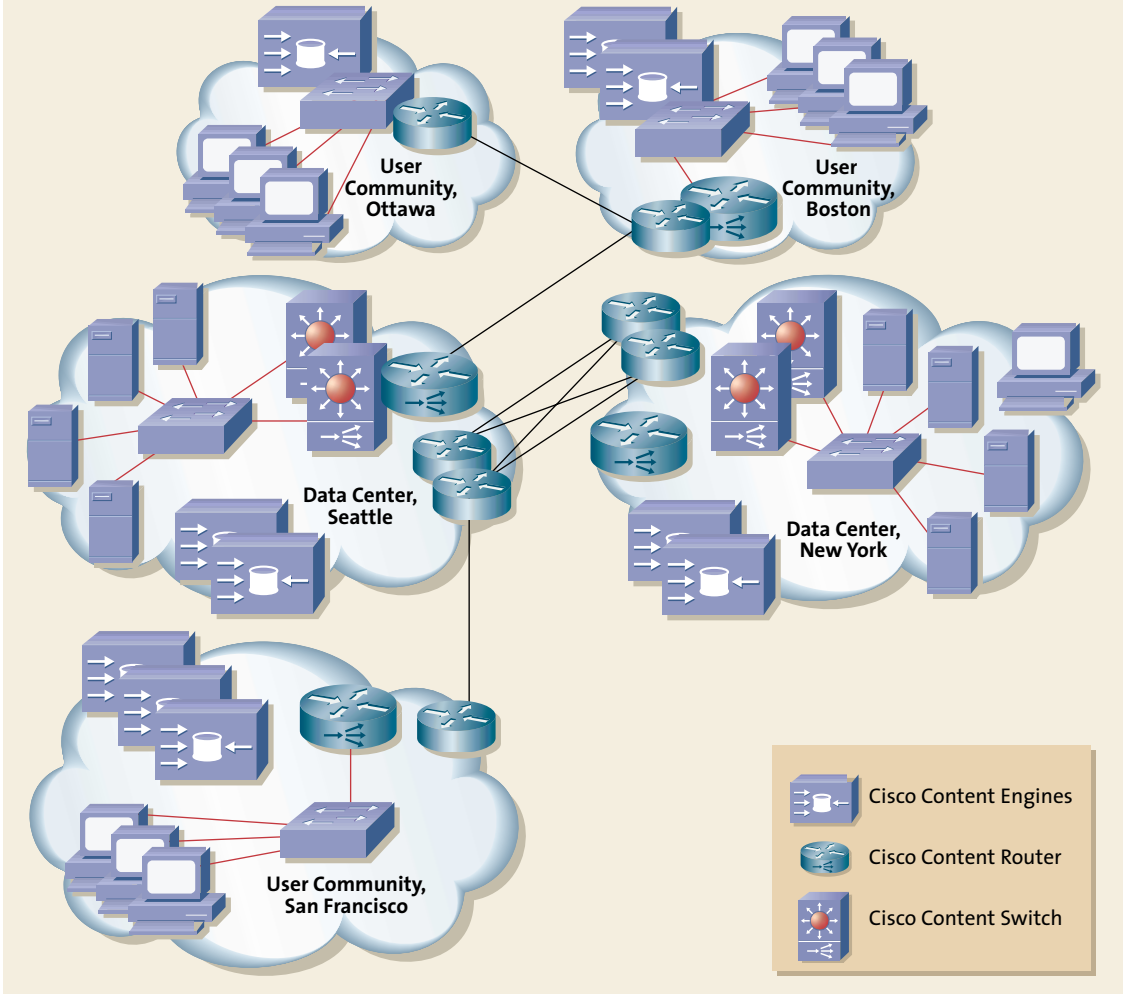
When designing content delivery networks, enterprises need to consider their distribution strategies. Where should the content engines—the edge nodes that store content for local delivery to customers—be located? How many content engines are needed at each site to manage the expected request and content-serving load? How much storage space is required to "hold" content replicas? How far to the edge does the content need to go? What quality is acceptable? And what are the cost tradeoffs of these decisions?

"Expect traffic growth that may be two to three times that of what you originally planned," says Mauricio Arregoces, a network design engineer on Cisco's Content Networking design team. "When designing a content delivery network, be sure to place your content engines away from network congestion points so as to avoid potential bottlenecks. For instance, if you have a highly trafficked T1 line between two cities, place the content engines behind the T1 so that requests and content are not delivered across the T1." (For more of Arregoces's design tips, see the sidebar, "Best Practices for Designing a Content Delivery Network" on page 19.)

An enterprise such as a financial house whose goal is to offer online learning and company communications to brokers nationwide through its intranet might opt for a more extensive distribution strategy to achieve the highest-quality Web and rich media content delivery. Imagine, for example, that a financial services firm has already built a nationwide, end-to-end Cisco network in North America, with data centers in Seattle and New York. The company has installed Cisco 7500 routers in each data center along with its Web-hosting equipment, including Web servers, Cisco Secure PIX™ Firewalls, intrusion-detection systems, Cisco CSS 11000 Content Switches, and other equipment. The Cisco 7500s are connected across leased DS3 lines.

Until recently, the company's brokers and analysts throughout the country have accessed one of these hub facilities to retrieve content. But as the firm has moved aggressively to provide online learning to rapidly improve broker productivity, the existing setup has proven to be insufficient. The problem is that many of

COMPONENTS FOR CONTENT DELIVERY NETWORKS



BRINGING CONTENT CLOSER: A Cisco content delivery network distributes content automatically to a number of remotely administered locations. These remote locations may be enterprise branch offices, content provider POPs, cable headends, or other "lights-out" facilities. When users request rich media content, the content delivery network determines which Cisco content engine is topologically closest to the requester and directs the requester to that ideal location.

the users are topologically remote from any of the sites and encounter multiple bottlenecks brought about by low-bandwidth links and propagation delays, resulting in stop-and-start, fuzzy video images.

The company opts to build a content delivery network. In the data centers and selected remote offices in San Francisco, Boston, and Ottawa, it installs the core pieces of the content delivery network, which include the following:

- **Cisco Content Distribution Manager** (CDM)—This platform incorporates technology from Cisco's SightPath acquisition to facilitate the provisioning, accounting, and billing of the services provided by the content delivery network. In addition, it provides transparent HTTP redirection of user requests to the best available content engine location. This product is only installed at the New York data center.
- **Cisco content engines**—The new Cisco Content Engine 500 series products are the next generation of Cisco Cache Engines, which are network appliances that transparently cache content from network servers and host it for content delivery networks. The Cisco content engines are placed at remote sites as well as at data centers. These products differ from the first-generation cache engines in that they can also download content from a central location and store it for local delivery to users, while constantly communicating with the CDM as to their operational status.
- **Cisco content switches**—These devices select the optimal content engine when more than one engine is present, and they intelligently load-balance traffic across content engines based on the availability of the content and load on each engine. Cisco content switches also monitor the health of the Cisco content engines to identify any problems and balance traffic around them if necessary.
- **Cisco content routers**—Devices such as the Cisco CR4400 content router are used to enhance the process of selecting the best available location for serving requested content. Content routers intercept

and operate on both Domain Name System (DNS) and HTTP processing in the network to improve the accuracy and reliability of optimal content placement.

Where to Move the Content? And When?

In the first phase of deploying a content delivery network, the enterprise uses the Cisco Content Distribution Manager (CDM) 4630/4650 platform to set policy, determining which content is to be moved to which edge nodes, the date and time, and also indicating for which LAN that edge-node will offer its content. For instance, if the enterprise wanted to provide a highly specialized training program on a class of mutual funds sold largely through brokers in the northeast region of the US, the financial services firm might select the content engines in the Boston area and indicate the time to replicate the content. The CDM would then index and store this information along with policies covering time for delivery, distribution scope, and bandwidth allocation for delivery. In addition to maintaining this information, the CDM would keep track of the underlined request activity (for the purpose of accounting and billing) by gathering the logs for all transactions in the content delivery network. This information may then be used for capacity planning and internal billing.

Now Showing on a Server Near You

Once policies have been set, the next phase of deployment is to distribute the content by replicating it to content engines at preselected locations.

Two mechanisms can initiate this action: the CDM in New York can push real-time file transfers out to the distribution points, or the content engines can pull the content from the CDM based on policies they receive from the CDM. Cisco's content switches can also replicate hot content within a site to relieve the overloading brought on by sudden demand, such as "flash crowds."

When content has been replicated, the content delivery network is ready. When a broker in Ottawa logs on to catch a half hour of her online training, her browser heads toward the Web-hosting equipment in New York. The CDM takes the broker's request for specific content and redirects the request to the best site on the global network, basing the routing decision on a set of metrics such as delay, topological proximity, server load, and a set of policies such as location of content that have been set up in the CDM. In this situation, the optimal content engines could be located in the Ottawa office, which is physically closest. But if the link to the office is congested, or the Ottawa server is overloaded, the request could be directed to the Boston office or

Continued

Best Practices for Designing a Content Delivery Network

Mauricio Arregoces (CCIE™ No. 3285), a network design engineer on Cisco's content networking design team, recommends the following practices for enterprise network design. From 14 years in the field, Arregoces boasts a wealth of experience in network design. "There are some important considerations that every designer should make note of early on," he observes. The following are his design recommendations for content delivery networks:

1. Design your network for growth. Expansion of your content delivery network enables the elimination of upstream bottlenecks and their effects on response time. As content is distributed closer to the end users, these users can more easily fill the pipe between them and the content. In general, the demand placed on content delivery networks between the content engine and user typically increases to two or three times that of the original experience. Make sure that you're able to scale the content delivery network as quickly as the new load demands are growing.

2. Place content engines away from bottlenecks. Always determine the current network bottlenecks and place the content engines on the other side of the contention points. In the example on page 17, for instance, if network usage patterns showed a heavily congested T1 line between New York and Boston, it would be wise to place an adequate number of content engines in the Boston area behind the T1 so that both the requests and the content would not need to be delivered over the congested line.

3. Use caution when planning replication. Choose to replicate content so as not to conflict with other important traffic. For example, you wouldn't want to schedule content replication of high-volume audio/video training material at the same time that you schedule company-wide backups and database system synchronization.

4. Be prepared to support more content. Content delivery networks become popular very quickly. Once yours is deployed, you will most likely be asked to support much more than what was originally planned. If you were asked to support one e-learning channel and a couple of corporate e-news channels on day one, expect that in the near future you'll be asked to support an entire suite of e-learning and e-news channels. This demand for rich content will no longer be limited by the peak bandwidth available within the core of the network, so end users are able to realize the full potential of broadband access, switched campus LANs, and branch office LANs.

5. Always monitor the content delivery network. You'll be surprised to learn how quickly new traffic patterns and new sets of applications will emerge. In a very short while, you'll be adapting the architecture that you initially developed for your content delivery network to accommodate more applications and richer content.

even to San Francisco if the metrics indicate that it is the optimal location.

The request will be routed using various pieces of information in the request. This information might include a domain name or an actual URL, the address of the user who requested the information, the user's location, or other information. After the request is routed, the user's browser connects to the nearest content engine.

Fine-Tuning the Images

When requests for content arrive at a specific site, the network must determine the best content engine to handle a request. The content switch selects the best server or cache and intelligently load-balances traffic across servers and caches, basing decisions on the availability of the content and load on the server or cache. Content switches also monitor the health of the content engines and server and help to identify and appropriately route around any problems.

Content switches are content-aware, so they have the intelligence to monitor server and cache farms beyond simple device availability, checking both service availability and content consistency. For instance, if an application has failed but the server is still running, the content switch would detect the fault by probing the appropriate TCP/UDP port, URL, or the size of a Web page object. When probing a URL, for instance, if the HTTP return code is not what is expected, the content switch would take the appropriate preconfigured action. This response might include placing the server off line and rerouting additional requests to an alternate site or server.

Cisco content switches also help to accelerate content delivery by enabling advanced performance and scalability of the server or cache farm. A content-aware switch has the ability to perform switching decisions based on information at Layers 4 and above—in other words, beyond IP address, protocol, and port numbers. This ability means that an enterprise can partition a site for greater performance by placing all the CPU-intensive portions of Web pages, such as scripts, in a server farm while placing all the HTML objects in another server farm. IT staff can then optimize each server farm to serve the type of content it supports, and the Cisco content switch can direct requests to the proper server farm based on the information in the URL. This capability can significantly enhance the end user's experience.

Content Networking Services

Ultimately, all of the content delivery network devices feed into a Cisco infrastructure that provides intelligent services at Layers 2 and 3 as well as security. Intelligent services at Layer 2 include private virtual LANs (PVLANs), enhancements to the Spanning Tree protocol (including the PortFast, UplinkFast, and Root Guard features), channeling capabilities using Fast Ethernet and Gigabit Ethernet, and VLAN access control lists (ACLs). Services at Layer 3 include advanced routing protocols, Hot Standby Routing Protocol (HSRP), router ACLs, and advanced policy and quality-of-service (QoS) mechanisms.

Security services are of utmost importance to distributed content, because each site maintains a copy of content from the original server. For that reason, security considerations applied to the original server site, including firewalls, intrusion-detection systems, router ACLs, and VLAN ACLs, should also be applied at distributed sites.

> By localizing the repeated content transmission through the intranet—a common performance bottleneck on networks—a content delivery network lowers propagation delays and latency and improves overall response time.

A New Era of Rich Media Content

A Cisco content delivery network endows intranets and the Internet with the Layer 7 intelligence that Cisco has traditionally applied to Layers 2 and 3. Through content-awareness, the content delivery network makes intelligent routing decisions based on content proximity and other topological considerations. The result is the realization of new, powerful applications that will have a significant impact on enterprise learning and employee, partner, and consumer communications. ▲▲

FURTHER READING

To learn more about content delivery networks, visit the URLs below.

- **cisco.com/go/cdn**
- **sightpath.com/solution_ent/products**