The Cisco Content Delivery Network Solution for the Enterprise

Executive Summary

The explosive growth of digital media has created an opportunity for enterprises to shift business strategies to incorporate new-world applications. Networks originated as a strategic way for enterprises to share information from one central location. Today enterprises are ready for the next-generation network—a way not only to share content, but also applications, rich media, and live events. This advance will affect business functions such as e-learning, e-communications, and e-commerce.

Content delivery is breakthrough technology allowing enterprises to distribute a new generation of scalable, accelerated, rich Web-based content, including TV-quality streaming media, for e-business and knowledge-sharing solutions. Content delivery replicates content to the "edge" of the network, minimizing the distance from the point at which content is requested and where it is served. Higher performance Web sites are the result. Content delivery overcomes inherent Internet network congestion issues, controls and maximizes bandwidth usage, and brings rich media services and applications to the desktop. Advantages of content delivery include central content control, efficient distribution to the servers at the "edge" of the network, and automatic routing of content requests to a local edge server.

Cisco offers the industry's only complete solution that combines five key ingredients for successful content delivery implementation: Distribution and management, routing, switching, edge delivery, and intelligent network services. Enterprises can improve network performance for delivery of on-demand content for both static and streaming media, greatly improve peak load site performance for live Web events and broadcasts, lower bandwidth costs, and enable new Web services using the products that comprise this solution.

Content Networking Overview

The Need for Rich Content in the Enterprise

Enterprises are demanding that networks accommodate increasing numbers of applications. E-learning, e-commerce, and e-communication will require greater bandwidth, higher reliability, and scalable networks. Pressure for these utilities on the Web will only increase over the next five years because they bring many benefits:

- Content control—Enterprises need the ability to control the entire distribution process—including time of day, bandwidth, and location—and want content refreshed quickly and easily through a simple-to-use graphical user interface (GUI).
- Cost saving—Cost of materials such as video cassettes and CD-ROMs can be reduced if content is distributed over the Web. Costs associated with travel can be reduced.
- Global reach—Enterprises can reach employees, customers, partners and students around the world, any time of day or night.

CISCO SYSTEMS

Consider the example of a corporation that uses video to train its workforce and keep client information up to date. Videotapes and CD-ROMs are expensive to produce and present logistical problems for the users. Though such an enterprise may look for ways to communicate through the Web to reduce the cost of videotapes and CD-ROMs, it may not find the solution with first-generation content-delivery technology. Sending video via e-mail or FTP is difficult to manage. Streaming over the Internet is not the quality experience desktop users want. High-impact training requires rich media files that can congest a company's network. First-generation technology lacks the sophistication demanded by high-profile clients.

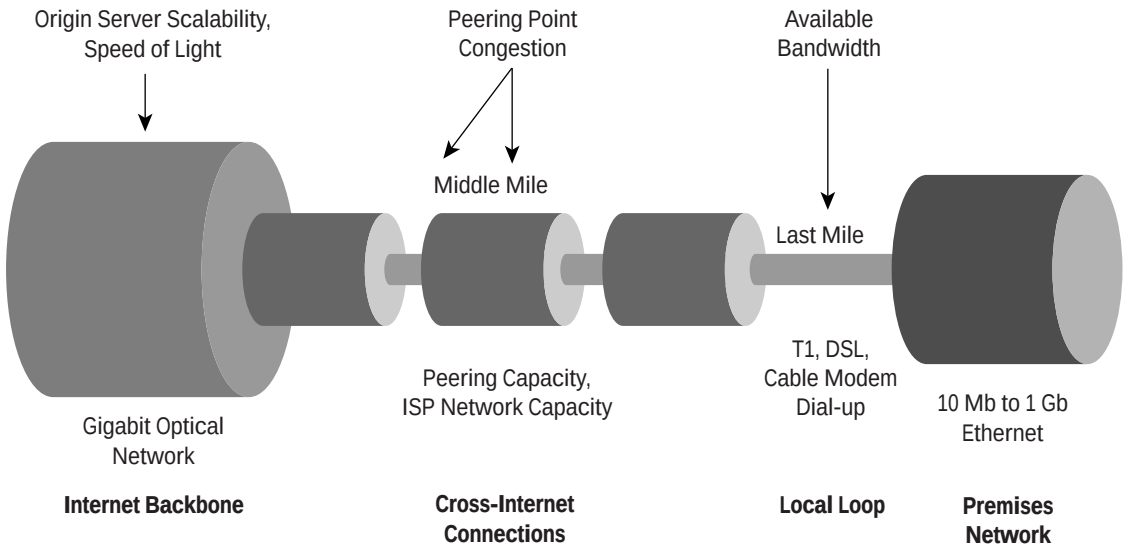
Barriers to Rich Content on the Enterprise

Web performance is a complex engineering topic. Today's Internet architecture is based on centralized servers delivering files to all points on the Internet. Bottlenecks can occur with this approach:

- Network link congestion—This is the data transmission delay associated with the capacity (bandwidth) of the telecommunication network. More users, richer media, and more complex applications consumes more bandwidth, resulting in increased congestion.
- Network equipment congestion—As network traffic increases, routers and switches must process more information packets. Packets collect in queues for processing. Overloaded network equipment discards queued data packets. They are then retransmitted, adding to traffic.
- Web server congestion—The larger the traffic flow on a Web site, the more requests for data must be served and processed by the server. Until the data is served, the requests sit in a queue, resulting in delay and a slow Web experience.
- Distance delay in the network—This is due to the time associated with data traveling over long distances.

Also, long distances in a network typically means that multiple organizations are associated with each step of the connection from server to browser. These intervening organizations may be outside the control of the site owner and primary carrier and may slow down the network.

Figure 1 Internet Congestion



Various methods have been employed to address these delays:

- Add bandwidth and faster network equipment—This is the simplest "brute force" approach, but can be expensive and only addresses bottlenecks that can be controlled.
- Geographic distribution of servers—This reduces congestion on both the server and network and distance delay through intelligent routing. Intelligent routing sends requests to an edge node that is close to the user so that content does not overload the server. Intelligent routing is limited, with no guarantee that content will be served from an appropriate server.

Traditional scalability methods simply are inadequate because general network and bandwidth gains continue to be consumed by Web traffic growth.

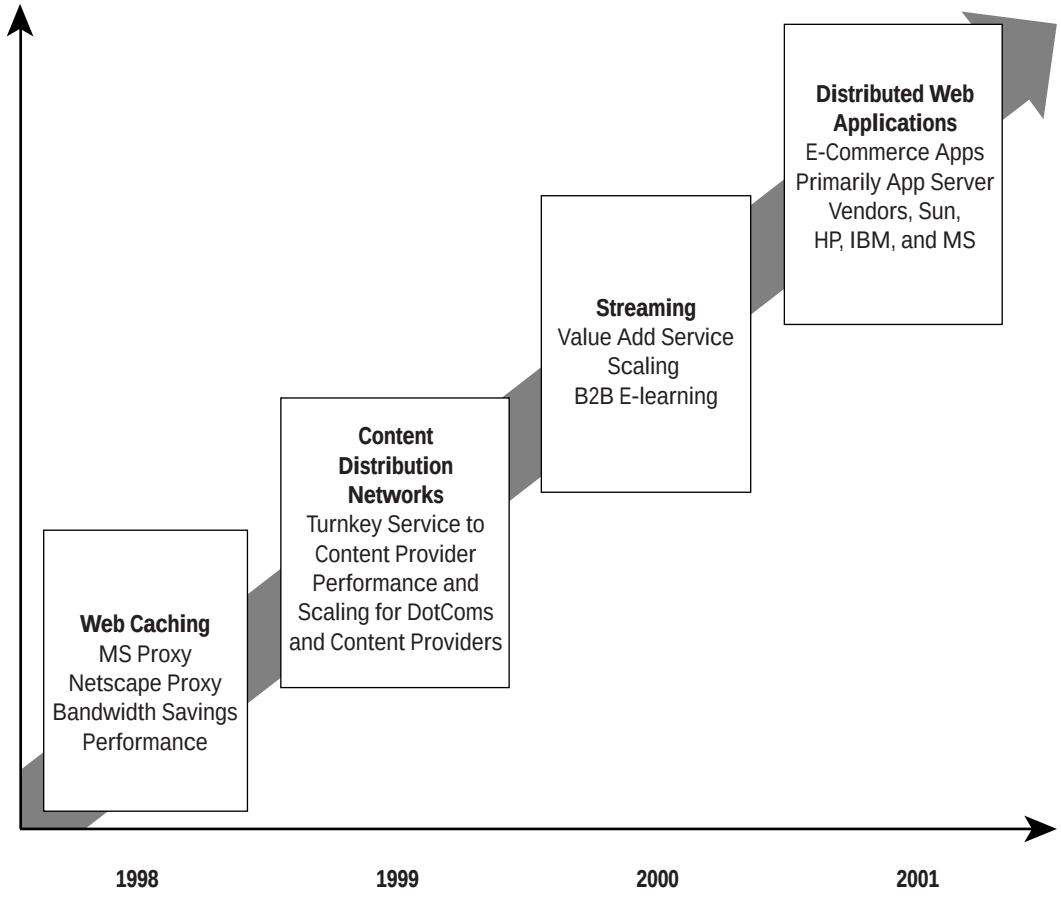
Content Networking Defined

"Over the next two years, a whole new layer of infrastructure that Forrester calls 'content routing' will emerge. This will optimize content delivery by linking applications with distributed network servers, and distributing requests across the network to maximize user performance."—Forrester Research, Scaling Web Performance, March 1999

Content networking is becoming mainstream technology because it drastically increases enterprise network performance for rich media at lower cost than traditional Web scaling. This reliable technology is becoming essential for mission-critical e-business applications. Cisco offers content delivery network (CDN) solutions that speed up Web performance and overcome issues associated with current delivery methods. Content networking provides for:

- Distribution of content to the edge of the network—Content delivery places servers throughout the Internet, close to where content requests originate. This principle mirrors some of the functionality of network caches, but with distinct differences. Content can be replicated to the servers where it can be served multiple times instead of being pulled by a user request. Network bandwidth is consumed only once, regardless of the number of times the content is served. Content will be available faster because the server has been pre-populated with data. Combining this feature with on-demand populated caching creates an integrated solution.
- Redirection of content requests to a local source—When a browser requests a URL, content delivery intelligently determines a local delivery server based on:
  - Geographical and network location
  - Presence of content
  - Current status of server (both availability and load)
  - Current status of network (traffic load and network errors)
  - Central control of content and network—All content and network devices are controlled from a central location, eliminating costs associated with remote administration. Content delivered by a content delivery solution has a single URL associated with it, regardless of its server source. A single URL eliminates the need to maintain multiple sets of content and Web sites.

Figure 2 CDN Evolution



Selecting an Enterprise CDN

Selecting a complete CDN solution requires that enterprises evaluate several aspects of the solution.

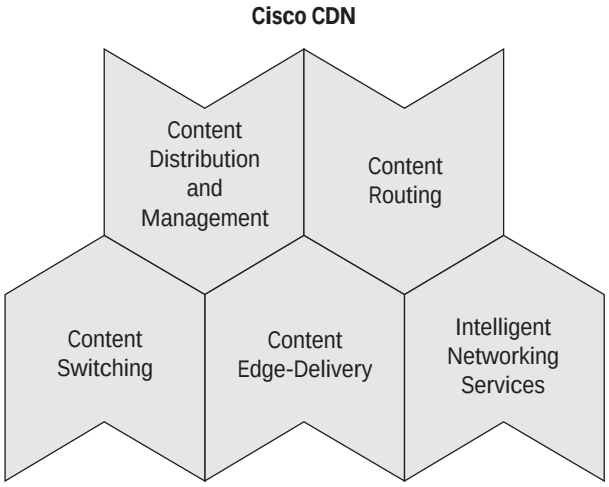
- Flexibility—Corporations use a variety of file types for rich media, including Shockwave, Flash, Director, SMIL, MPEG, JPEG, etc. A corporation may use a combination of these file types in everyday operations. An enterprise should not be limited by its networking equipment. Content creators ease limitations imposed by network capabilities when they deploy a CDN that is file format agnostic.
- Ease of use and administration—CDNs should be transparent to desktop end users and should not require special browser plug-ins or configurations. Global coordination of network elements allows uniform access, regardless of a user's location or the network's topology. CDNs should be easy to administer and maintain, in spite of the fact that, as often happens, CDN components may be dispersed in multiple office locations.
- Scalability and future proof—Networking needs constantly change. The CDN solution should be designed with future needs in mind. A scalable CDN gives enterprises the ability to meet the needs of today's environment and protect against future expensive, disruptive forklift upgrades.
- Cost of deployment and ownership—The workload on network administrators is drastically reduced by central administration of content and CDN devices. Cost of ownership is thereby reduced. The key to choosing a CDN solution is to pick one that enhances the current network, rather than replaces it entirely at substantial cost.

Elements of Content Networking

Five key elements comprise content delivery networks:

- Content Distribution and Management—Configures CDN devices, manages content, assigns network policies, and monitors network performance
- Content Routing—Reliably redirects user requests to the best location in the CDN
- Content Switching—Intelligently switches and balances traffic load across multiple servers
- Content Edge-Delivery—Delivers content from the network edge to the desktop
- Intelligent Network Services—Provides network services such as security, QoS and Multicast in the network infrastructure

Figure 3 CDN Key Elements



Content Distribution and Management

Internet and intranet traffic grows at a phenomenal rate—doubling every 100 days. Rapid increase in network traffic creates networking challenges, including:

- Wide-area network (WAN) bandwidth congestion and high transmission costs
- Maximizing network service quality
- Maximizing and controlling the availability of Internet and intranet content as seen by clients
- Cost-efficient network scalability

With the Cisco content-delivery network solution, enterprises eliminate these challenges by controlling and managing the distribution process.

Content Distribution

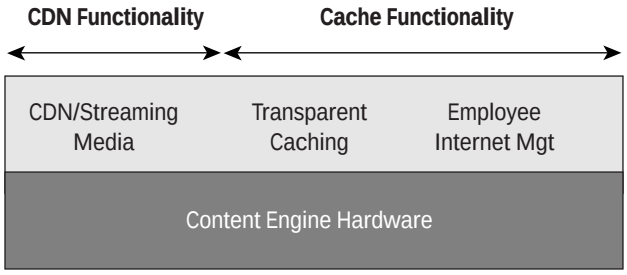
Solution—Localize Traffic Patterns with the Cisco Content Engine

The most efficient solution to these networking problems is to use the existing network to enable content requests to be fulfilled locally. This solution addresses networking challenges in the following ways:

- Accelerated content delivery—Traffic localization accelerates content delivery by locally fulfilling content requests rather than traversing the Internet and intranet to a distant server farm. This solution helps protect the network from uncontrollable bottlenecks, delivering more consistent network service quality and content availability.
- Optimized WAN bandwidth usage—Traffic localization minimizes redundant network traffic that traverses WAN links. As a result, WAN bandwidth costs either decrease or grow less quickly. This bandwidth optimization increases network capacity for additional users and new services.
- Greater access control—Traffic localization enables administrators to monitor and filter access to objectionable content. The result is more productive WAN bandwidth usage and lower enterprise legal liability.

Traffic localization is a traffic engineering problem because it requires the network intelligence to optimize traffic flows based on specified parameters. Therefore, the first step in building a traffic localization solution is to ensure that the existing network supports this capability. This capability can be achieved by enabling content routing technology, such as Cisco IOS® software's Web cache communication protocol (WCCP), at key points within the network.

Figure 4 Content Engine



Once the network is in place, Cisco Content Engines are added strategically within the network to complete the traffic localization solution. Content engines provide transparent caching and/or CDN functionality to store frequently accessed content and locally fulfill successive requests for the same content, eliminating repetitive transmission of identical content over x links.

Figure 5 Caching and CDN

Capability	Caching	CDN
Content Source	All Content Requested	Only Assigned Domains and Nes
Content Population	Pull or Cache-driven Pre-population	Centralized Pre-population, Based on Policies
Request Routing	Local - Intercept Traffic and Deliver from Cache	Global - Redirect Request to Optimal Node
Load Balancing	Local Clustering	Global, Across Entire CDN
Content Management	None Needed - Automated	Centralized Content Management
Content Retention	Store Frequently Requested Files	Policy-driven. Can Ensure 100% Hit Rate

Integrating network content engines results in a delivery solution with a low cost of ownership, enabling enterprises to deploy content on a wide scale cost effectively while gaining benefits throughout the entire network. Network-integrated content engines have three basic properties in common:

- Managed like networking equipment, resulting in minimized operational costs
- Designed like high-density networking hardware, resulting in better physical integration into the network infrastructure as network extensions and minimizing costs associated with leasing rack space
- Transparently inserted into the network, resulting in minimized deployment and operational costs and greater content availability

Content Replication

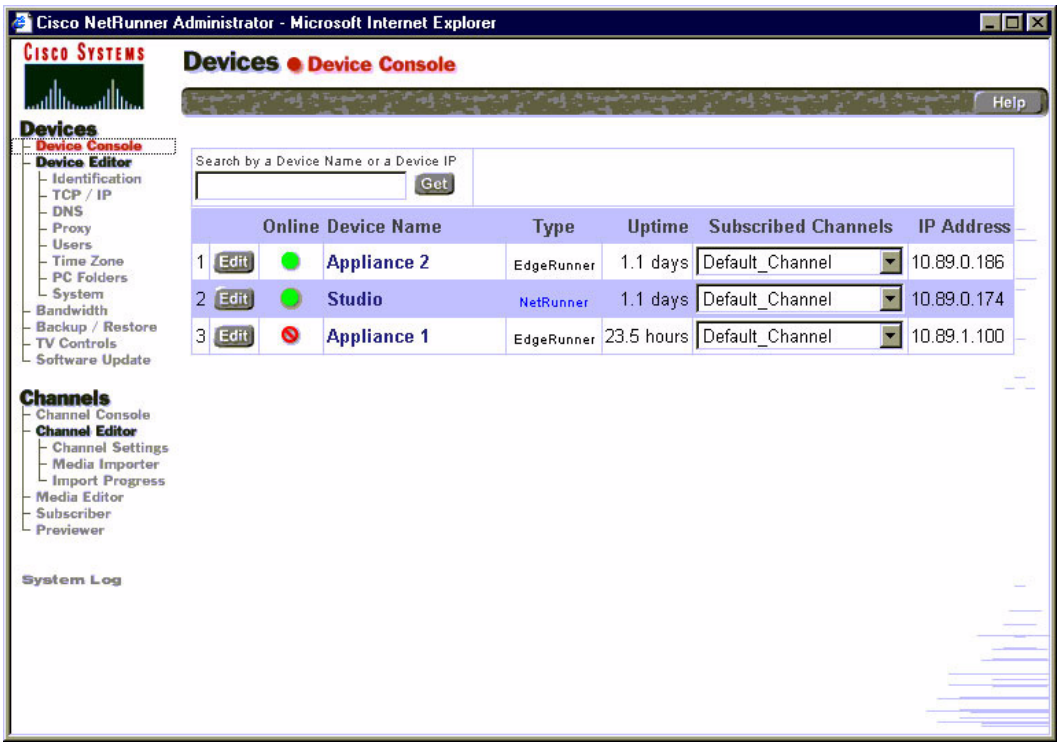
Cisco provides a minimally-intrusive technology that directs user requests to the optimal point in the network. Flexible architecture means the user may delegate the Web site domain to the content-delivery system, or perform a simple edit to the URL domain name to enable content to be served from the same system.

Entire Web sites can be designated for delivery, or a subset of the content based on a user-defined set of rules (either just GIF delivery or all files three levels deep from http://www.cisco.com). Once the content has been flagged for delivery, no changes are required to the existing Web site. All user requests are routed to the central Web server, which then redirects the request to the optimal content engine. This system includes the use of a single Web site that captures all user and impression information while keeping maintenance costs down.

Content Management

At the heart of the Cisco content delivery network is Cisco software functioning as the content distribution manager (CDM). The content distribution manager provides complete control over the media distribution network—including all of the Cisco Content Engines located at end-user sites, which receive and stream media to audiences over the local-area network (LAN). Through the content distribution manager's Web browser-based user interface, enterprises can configure and monitor Cisco Content Engines anywhere in the world, import and preview any rich media—including TV-quality video—and generate media URLs for intranet and extranet Web sites. Enterprises can also set maximum bandwidth usage rates for distributing media to Cisco Content Engines over the WAN, as well as from content engines to end users over the LAN to eliminate network congestion.

Figure 6 CDM User Interface



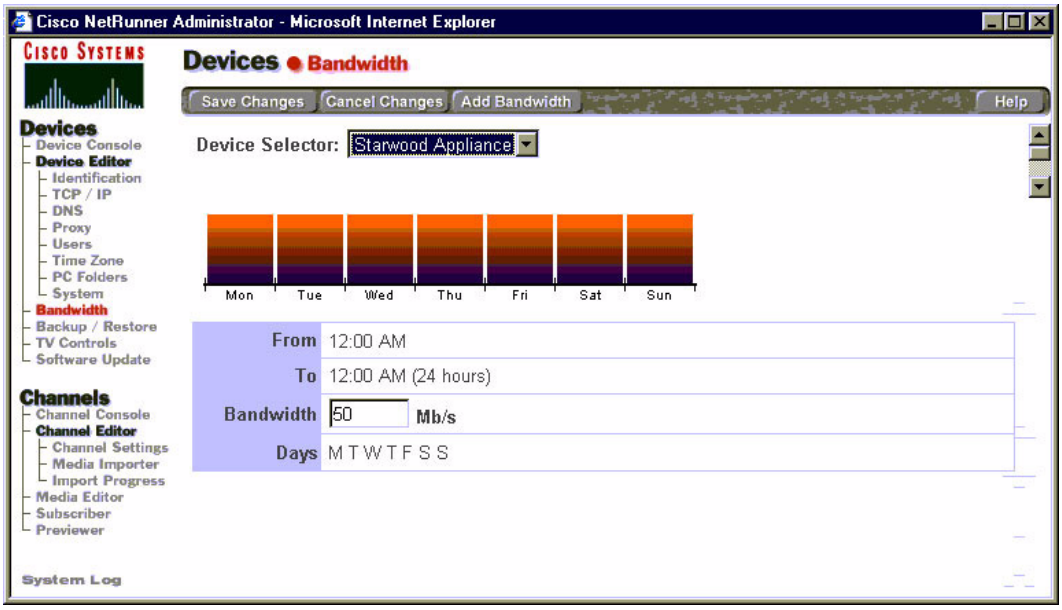
Total Control over Content Distribution

- Breakthrough SODA technology for cost-effective, fully-automated replication of high-bandwidth media files to Cisco Content Engines on any Internet-connected LAN
- Simple, intuitive commands for adding, removing, and selecting media files for replication
- Preview of media files prior to replication
- Central control over thousands of content engines
- Intelligent Routing and Server selection occurs automatically
- Secure and fault tolerant file transfer using SSL encryption for secure media transfers
- Channel configuration for media distribution to any number of discrete audiences using “distribution lists”

Complete Control over Network Bandwidth

To eliminate network congestion, the CDM provides enterprises with total bandwidth control. Organizations can set maximum bandwidth usage over the WAN from the content distribution manager to the remote content engines, as well as maximum LAN bandwidth usage from the content engines to end-user desktops. Large files can even be delivered overnight or during off-peak hours, and even specify replication by time of day and day of week to avoid impacts on network activities.

Figure 7 CDM Bandwidth Settings



Unsurpassed Web Integration

For fast, easy integration with the Web, the Cisco Content Distribution manager creates one URL per media file, which can be accessed by any number of content engines, regardless of how many are in the field. Cutting and pasting one URL into the existing intranet or extranet page gives enterprises a fast, simple way to manage and deliver content effectively without having to manage every transaction. To view media, end users simply visit the page and click on the new media link. The command is redirected to the nearest content engine, which streams the media locally, ensuring that the end user gets the best quality Web experience.

- Automatic generation of thumbnail reference images and sample Web pages for integration with corporate extranet, intranet, and Internet sites
- One URL per media file provides seamless integration into any Web site
- Integrates with standards for Web multimedia presentation including HTML/DHTML, XML, and SMIL

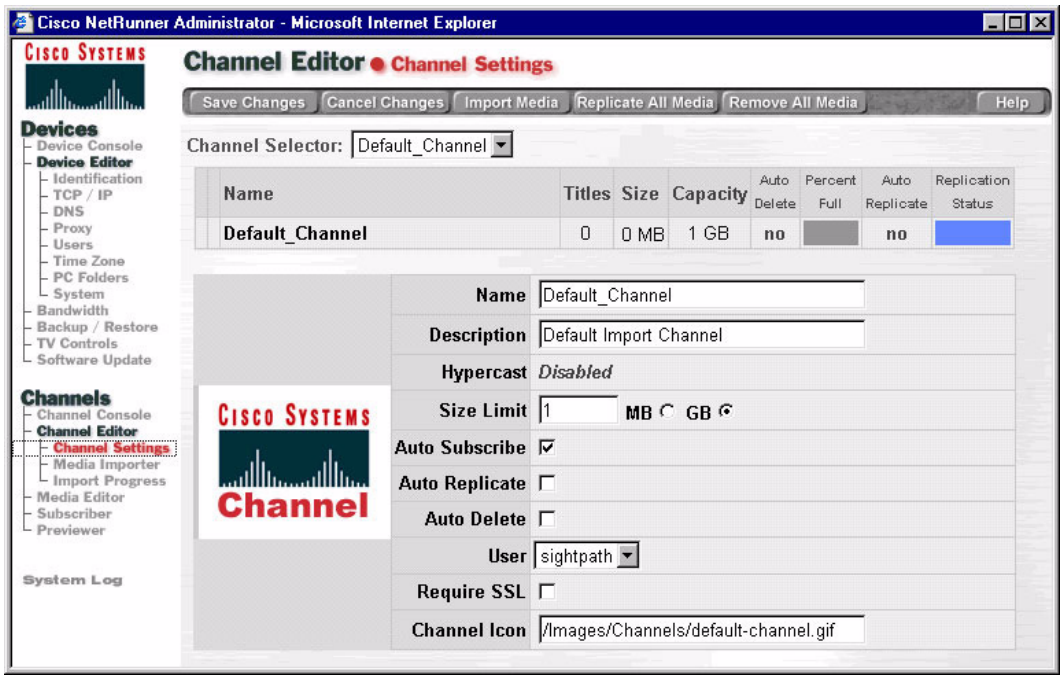
Reach Audiences Anywhere with Alternate Media Delivery

While the key benefit of the Cisco CDM system is the ability to deliver high-bandwidth content over any IP network, the content distribution manager also makes it possible to deliver rich media even if the target audience isn't connected to a Cisco Content Engine on a LAN. Using the content distribution manager's alternate media delivery feature, enterprises can create a "smart" media URL that plays alternate, lower-bandwidth media for users who are out of the office. Once viewers are back on the LAN, they have the rich content delivery experience. In the meantime they can still access content—whether it's low-bandwidth video, a custom HTML page, or a PowerPoint presentation—when they are on the road.

Fast, Secure Media Replication with Zero Incremental Bandwidth

The Cisco CDM uses replication to distribute its content to each content engine on the network. It also uses SSL encryption to safeguard sensitive data. Since each engine holds a replica of the original content, managing media is simple, even if an enterprise has thousands of content engines in the field. That's because the system allows the administrator to distribute or delete media from one source. Since the system distributes media using low bandwidth over the WAN, then delivers media to the desktop using the bandwidth available on the LAN, it's an efficient use of network resources. Whether media is requested once or a thousand times, there is no additional load on the WAN. Only Cisco offers this innovative, cost-effective approach to delivering any rich media. Using HTTP for file transfer, the content distribution manager's and content engine can communicate even behind the safety of a corporate firewall.

Figure 8 Channel Settings



Flexible Importing

- Import standard video formats, including MPEG-1, MPEG-2, RealVideo, Windows Media, AVI, QuickTime, and more
- Import any standard non-video file type, including PDF, Binary, CAD/CAM, PowerPoint, and Shockwave
- Provides three flexible importing options: FTP, drag and drop from Microsoft Windows Network Neighborhood, or copy from Web server

Fault Tolerance and Fault Prevention

The Cisco content delivery solution provides both device fault tolerance and networking fault tolerance, eliminating any single point of failure. If a content engine fails, traffic is automatically redistributed among other content engine cluster members and the network topology is rebuilt. In the event a failure occurs during a file replication, upon reconfiguration of the network the replication will resume at the point it left off.

Standards Based

Current technology allows a single Cisco Content Engine to deliver from 20 to 100 simultaneous streams of MPEG video and store up to 150 hours of media. Content engines use TCP/IP and HTTP for communications, and work with industry-standard LANs, PCs, Web browsers, and media players. Because the products use standards-based protocols, delivery of any file format to the desktop is easy and seamless. With no new client software required on the desktop, Cisco Content Engines make it easy to transform intranet or extranet Web sites into simple and affordable delivery mediums for any rich media. Quality communications have never been easier or more cost-effective.

Content Routing

The ability to redirect user requests to the best content router is an integral piece of the CDN solution. Content routers and content engines are in constant communication with each other using Self-Organizing Distributed Architecture (SODA) architecture. The CDN devices form a spanning tree, and the tree is updated when devices are added or removed from the network or when network conditions change. The content engines and content routers communicate using HTTP to build and evolve the network hierarchy based on dynamic network conditions. The SODA algorithm learns the network topology and performance improves with time and usage. When a browser selects content from a specific Web site, the CDN will look at the source IP address and redirect the request to the optimal content engine.

The content router will redirect a request to the serving content engine based on the following criteria:

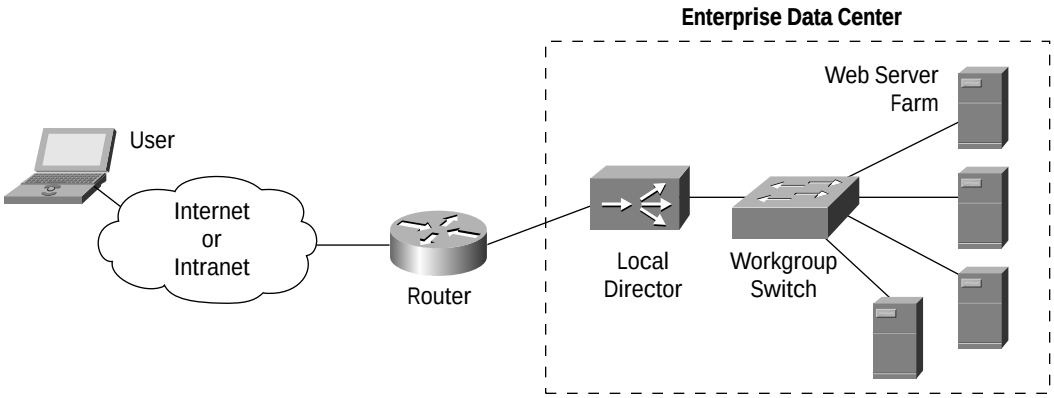
- Presence of Routed Content—Content engines subscribe to channels in the CDM. In order for a browser to be redirected to a content engine in the serving area using HTTP, the requested files must be available on the content engine. Content routing within the CDN resolves these requests.
- Geographical Location/originating network—In many cases a user's browser will be redirected to the closest content engine because the shortest distance to be traversed is often the best serving location. However, the global redirection function may dictate that a file is best served by a content engine outside of the local cluster.
- Network Status—Content engines and content routers will keep track of network conditions by exchanging test messages. Based on test results, content routers will update their spanning tree maps to reflect network conditions.
- Current Content Engine Health—When a content router or content engine fails it will be removed from the network topology and will not service requests. When a content router or content engine joins the CDN, it is automatically placed into the hierarchy.
- Current Content Engine Load—Content routers will not send HTTP redirects to overloaded content engines. When the load on the content engine decreases, it will automatically become available and begin servicing new requests.

Content Switching

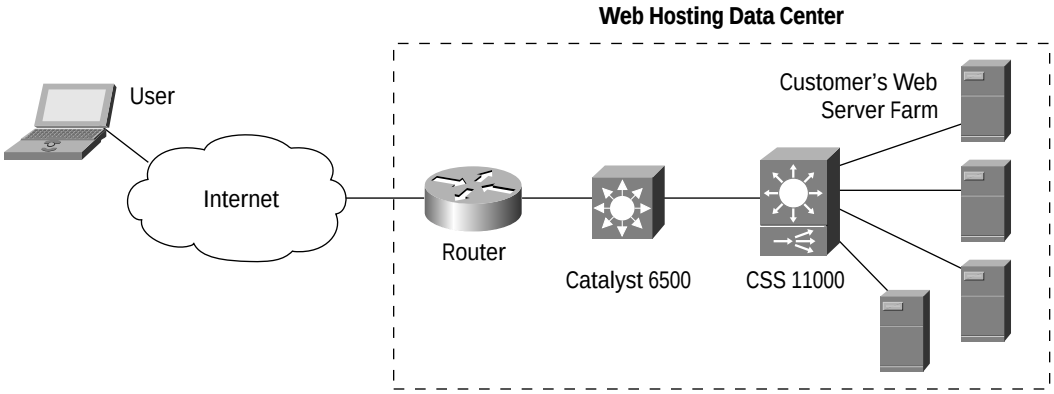
The growing demand for mission-critical Transmission Control Protocol/Internet Protocol (TCP/IP)-based application services such as e-commerce, video, enterprise resource planning (ERP), and e-mail is motivating companies to increase the reliability of their network and server systems. The Cisco LocalDirector series and the CSS 11000 Content Service Switches are critical components of reliable and integrated network and server systems. The Cisco LocalDirector series offers enterprise customers and Internet service providers (ISPs) a network-based server and application clustering solution. It is an integrated hardware and software solution that intelligently balances the load of user traffic across multiple TCP/IP application servers (Figure 1). All physical servers appear as one virtual server, with the result that only a single IP address and a single Universal Resource Locator (URL) is required for an entire server farm. The CSS 11000 series of Content Services Switches extend the capabilities of enterprise and e-commerce customers with a combination of high speed switching and intelligent routing of Web requests to the best site or server. Each request is routed based on URL, which identifies content being requested; cookie, which identifies session and user priority; browser type, which identifies access device—PC, Web phone, PDA; and language preference.

Cisco LocalDirector assures continuous availability of content and applications with proven techniques for actively managing servers and connections in a distributed environment. By distributing user requests across a cluster of servers, Cisco LocalDirector optimizes response and system capacity, and reduces the cost of providing large-scale Internet, database, and application services. Its integrated security device also protects servers from unauthorized access.

Figure 9 Content Switching in the Enterprise



Content Switching in the Enterprise



Content Switching for Enterprise and E-commerce

Cisco LocalDirector is tightly integrated with a Cisco network. The Accelerated Server Load Balancing (ASLB) feature works in conjunction with the Catalyst® 6000 and Catalyst 6500 to accelerate TCP sessions. The ASLB feature set ensures load-balancing performance greater than 800 Mbps and up to 15 Mbps. The multi-node feature improves availability. The Cisco LocalDirector software, in conjunction with Cisco routers, now allows multiple points of access to the servers—the ultimate in availability. The Cisco quality-of-service (QoS) Policy Manager manages Cisco LocalDirector QoS attributes.

The CSS 11000 series works with Catalyst switches and Cisco routers to enable content services in enterprise data centers. It also works for enterprise and e-commerce customers who use a Web hosting service provider. The CSS 11000 series enable comprehensive Web site security. Denial of service attacks are eliminated. Web transactions are filtered through Content Access Control Lists. Firewall load balancing enables scalability for Cisco PIX™ Firewalls, providing high performance and eliminating potential single points of failure.

The CSS 11000 enables e-business applications with transaction assurance, providing "sticky" connections based on cookies and SSL session ID. Users stay connected to the right server through transitions from SSL to non-secure HTTP, eliminating dropped shopping carts. User transactions can be prioritized, or routed to specific servers based on cookies. Users can be routed to the correct sites or servers for the access devices they are using (PDA/Webphone) or the language they prefer.

Integrated content staging and replication combined with the intelligence to understand the frequency of requests for particular content allows the CSS 11000 to dynamically replicate hot content to overflow servers or caches. It also facilitates content publishing and freshness capabilities by replicating specific content to multiple subscribers based on changed content, time of day, or explicit command using CLI, SNMP, or XML API from a server-side application.

The architecture of the CSS 11000 is based on a scalable hardware platform ranging from 8 10/100 ports in a fixed configuration to a modular chassis with up to 64 10/100 and 32 Gigabit ethernet ports. The high speed switching is enabled through per port processing of established flows while the content intelligence and flow-classification are handled by a multi-processor control plane. This ensures all control plane resources are available to traffic at any port with the performance to support the busiest Web sites. Once mapped, flows are switched directly from server to client at wire speed.

Figure 10 Content Switching for Enterprise and E-commerce

High-Availability Feature Sets

- Multi-node architecture increases data center online time because it builds server systems with multiple points of access.
- A fail-over mechanism eliminates all points of failure, enabling robust, high-availability server farms.
- Automatic detection and removal of failed applications improves overall server farm management and reliability.
- The Cisco LocalDirector Content Verification System monitors servers for application availability, application health, and database connectivity, effectively routing and directing traffic only to available applications.
- A hot-standby server allows servers or ports to back up other services and ports, increasing availability.
- The hot-standby configuration replication feature reduces administrative startup costs by automatically transferring the primary LocalDirector configuration to the secondary one.
- The CSS 11000 series enables content, site, server, and application redundancy.
- Dynamic Content Replication enables Flash Crowd Insurance—the distribution of hot content to overflow servers and caches in response to a flash crowd, dynamically scaling the Web site in real time.

Scalability Feature Sets

- A real-time embedded operating system ensures the highest performance, with models capable of 80- or 400-Mbps throughput.
- The Cisco LocalDirector supports the highest-traffic Internet sites, with capacity to support 7000 connections per second or 30,000 connections per second.
- ALSB with the Catalyst 6000 increases total load balancing throughput to over 800 Mbps throughput and packet forwarding up 15 Mbps.
- 384 MB of RAM ensures high performance while operating Layer 7 services such as SSL Sticky.
- Session distribution algorithm supports more than one-million simultaneous TCP sessions to accommodate rapid user population growth.
- Multi-node features allow for incremental scalability of load-balancing devices, improving return on investment.
- The CSS 11000 series combines a high-speed forwarding plane with a multi-processor control plane to deliver the highest performance for HTTP transaction switching and routing. ZD labs validated speeds up to 130,000 HTTP transactions per second, or more than 11 billion per day.
- Because it is compatible with any server operating system, administrators can mix and match server hardware and operating systems while protecting IS investments.

Server Connection Management Feature Sets

- Support for all common TCP/IP Internet services—Web, File Transfer Protocol (FTP), Telnet, DNS, and Simple Mail Transfer Protocol (SMTP)
- Simple addition and removal of servers allows for an easy upgrade of applications servers. The ability to scale a server farm and direct traffic based on application allows configuration flexibility.
- Packet coloring allows implementation of powerful QoS networking.
- SSL Sticky provides the ability to maintain client-to-server persistence based on SSL session identification, improving the overall Internet browsing experience.
- Cookie Sticky provides the ability to maintain client-to-server persistence based on Cisco LocalDirector or server-generated cookies.
- Cisco LocalDirector ensures that license counts or applications are not overwhelmed by setting a maximum connection.
- The CSS 11000 series can examine every request in an HTTP 1.1 persistent connection and open new server connections as necessary for content located on different servers.

Server Security Feature Sets

- The NAT function effectively prevents direct access to real servers through conservation of IP addresses and utilization with unregistered IP addresses.
- Cisco LocalDirector offers robust traffic filtering based on source IP address or service.
- The CSS 11000 series provides integrated denial-of-service protection, eliminating incomplete or malformed TCP and HTTP connections.
- Content Access Control Lists eliminate unwanted traffic based on network address or URL.
- Firewall Load Balancing provides high performance firewall clusters and eliminates single points of failure.

Load-balancing and Content Switching solutions in general and Cisco LocalDirector and the CSS 11000 series in particular are successful because they provide five fundamental benefits in server-farm environments. These benefits are:

- The ability to manage, scale, and reduce the variability of traffic loads.
- A low-cost, easy-to-implement, high-availability strategy for managing server traffic.
- An ability to intelligently manage connections between clients and servers.
- An integrated and effective security system to protect data and applications on servers.
- The ability to provision content and user services for e-businesses.

Content Edge-Delivery

Although CDNs are designed to handle any file type, streaming media is the most obvious application to take advantage of this new network for enterprise applications. Corporations want to use their networks for corporate communication and training, but many networks are not designed to handle video. Many enterprises can handle a 1.5 Mb/s MPEG1 or 500 Kb/s RealNetworks file across a T1 connection, but bottlenecks often result. CDNs address the issue by pushing rich media content to the network edges. The Cisco Content Engines at the edges of the network contain streaming media servers that take advantage of the CDN switching, transparent redirection and management to stream video across the LAN. The content distribution manager controls all server policies. Content engines have the ability to stream multiple file formats from different servers on the same content engine. Each of these servers offers the ability to stream live or VOD to the end user desktop.

SODA

Self-Organizing Distributed Architecture (SODA) is the patented technology of the Cisco content delivery solution. The Cisco solution—a single source for technology in the areas of caching, content distribution and management, load balancing, and redirection—is unique. Other suppliers can provide some of the technology, but only Cisco offers end-to-end advanced content delivery solutions.

Self-Organizing Content Engines

Cisco Content Engine nodes automatically organize themselves into a single system that can be expanded simply by adding more nodes. A new node automatically configures itself based on current network topology and performance. Its place in the network is optimized around bandwidth, network congestion, and geographical location, as well as its relationship to other nodes.

Each node is part of a sophisticated spanning tree routing hierarchy that efficiently handles replication of published content. The note splits live and published content and serves both. Each node monitors the current state of the network and its peer nodes to adjust to changing conditions: It understands when nodes have dropped out of the network, whether the network is congested, or if a node is busy. Changes in conditions can be detected nearly instantaneously. Corrections are made to prevent service interruptions.

Automated, Spanning Tree Replication

The Cisco content networking solution distributes content to the edge of the network by replication. Replication of both pushed and pulled content is transmitted through the network to remote content engine nodes at the edge. Since replication occurs from an origin server to a content engine node, and from a content engine node to other content engine nodes, efficient use of the downstream network bandwidth is assured. The spanning tree replication topology can keep bandwidth requirements at a central Web site to a minimum while utilizing less expensive edge bandwidth.

The replication technology requires no manual intervention to mirror critical Web content. Content is organized into distribution networks (usually associated with customers) and each network is subscribed to a set of content engine nodes. Content can subscribe either to a global set or a subset based on the target audience. Policies can be set to determine the right level of distribution to a specific Web site, whether by geographic region or by number of servers.

Single-URL Redirection

Cisco content networking is a minimally intrusive redirection technology. Flexible architecture means the user may delegate the Web site domain to the content delivery system or do a simple edit to the URL domain name to serve content from the same system.

Entire Web sites can be designated for delivery. A user may choose to deliver only a subset of the content—just GIF delivery, for example, or all files three levels deep from http://www.cisco.com, for another. Once the content has been flagged for delivery, no changes are required to the existing Web site. All user requests are routed to the central Web server, which redirects to the optimal content engine node. This system uses a single Web site that captures all user and impression information, keeping maintenance costs down.

Cisco Streaming Server (CSS)

The content engines contain a CSS that is equipped to stream any file format, including MPEG 1 and MPEG 2, using HTTP. These servers can stream any video and audio file.

RealNetworks

Enterprises requiring the advanced features of RealNetworks can take advantage of the RealNetworks G2 server contained in each of the content engines to stream RealNetworks files in their native RTP/RTSP formats. All management and configuration is handled by the CDM, and there is no local maintenance or configuration required at the content engines.

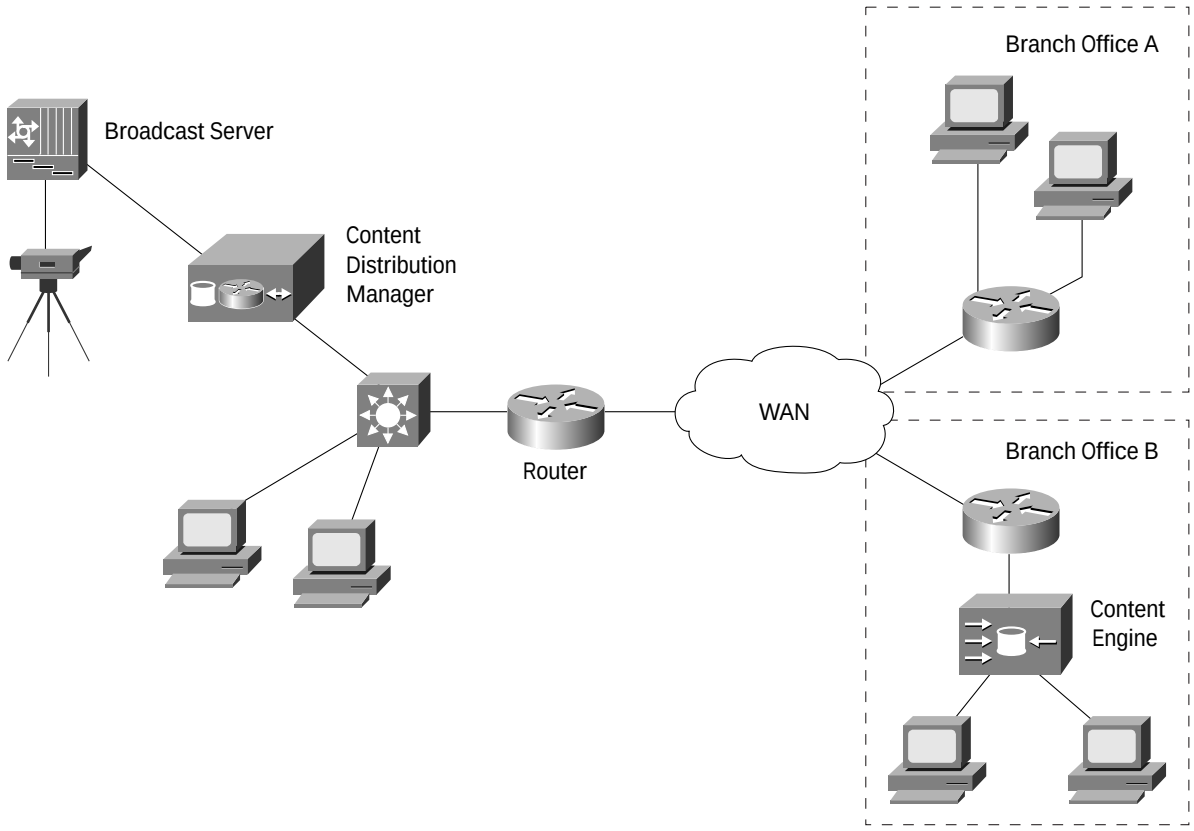
Live Streaming with IP Multicast

The Cisco IP/TV Broadcast Servers are ideal for multicasting live events or prerecorded programs on a scheduled basis. Source files can come from video cameras, VCRs, satellite feeds, cable feeds, or existing ASF, AVI, or MPEG digital formats. Delivery can be in real time without jeopardizing network bandwidth resources. Cisco IP/TV Broadcast Servers can multicast a live program and numerous scheduled prerecorded programs simultaneously. They deliver events as they happen—such as a CEO's address to thousands of people—using individual streams of network bandwidth. The Broadcast Servers leverage the multicast capability of Cisco routers and use SODA to push content to the local content engines for video-on-demand (VOD).

Two Broadcast Servers are provided, depending upon application requirements and available bandwidth.

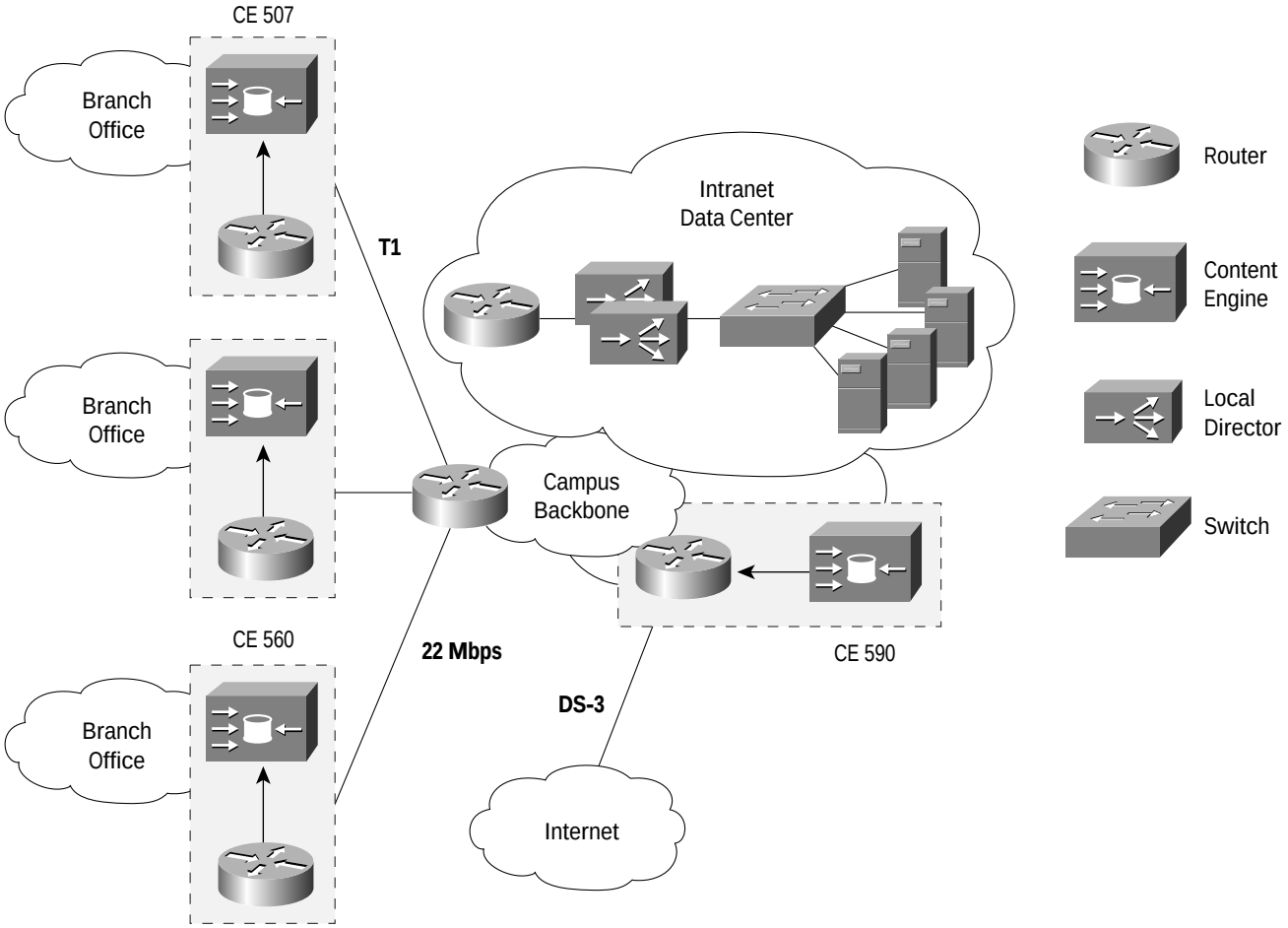
- The Cisco IP/TV 3422 Broadcast Server uses MPEG-4 video to stream real-time and prerecorded content. This server is appropriate for low-bandwidth connections to remote sites, to field offices, or over the Internet.
- The Cisco IP/TV 3423 Broadcast Server uses a wide variety of formats, including MPEG-1, MPEG-2, MPEG-4, Indeo, and H.261 compression techniques, to capture real-time and prerecorded content and stream it over a network. It also helps to balance quality with bandwidth limitations.

Figure 11 Live Streaming



- Transparent network caching and desktop delivery—WCCP-enabled router or content router redirects a user's request for Web content to an optimal content engine for local delivery across the LAN. Clients do not need to configure their browsers to point to a specific proxy server.
- Hierarchical deployment—Content engines can be placed in several network locations to manually or automatically build a spanning tree in a hierarchical fashion.
- Scalable clustering—In conjunction with WCCP-enabled routers or content routers, customers can scale performance to support higher WAN links, greater throughput, or storage by hot inserting additional content engines into the CDN.
- Fault tolerance/fail safety—If any content engine in a cluster fails, the cluster heals itself. The WCCP-enabled router or content router redistributes the failed engine's load evenly among the remaining content engines.
- Overload bypass—The content engine detects when it is overloaded, refuses requests, and forwards them to the origin Web servers or other content engines. The overload bypass on/off triggers are automatically determined by CPU and file system load.
- Dynamic client bypass—Some Web sites require clients to be authenticated using the client's IP address. However, when a network cache is inserted between a client and a Web server, the Web server only sees the cache's IP address and not the client's IP address. The content engine has the ability to detect this scenario and automatically send an HTTP retry to the browser. When the browser retries, the content engine refuses the next request and sends it directly to the Web server. Thus, the origin Web server will see the client's IP address, authenticate the client, and respond directly to the client.
- Reverse proxy—Content engines can serve as the front end of a server farm and handle a small number of high-demand URLs. When an incoming Web request reaches a WCCP-enabled router, the router maps the request to one of the content engines. The request is then served to the browser without ever connecting to the origin Web server.
- Support for streaming media—Requests for streaming media files can be redirected via http to a local content engine containing the cached or pre-loaded file for high bandwidth playback over the LAN to a user's desktop.
- Content aware—Content engines communicate with content engine peers outside of the local cluster to obtain content as it is needed.
- Self registering and organizing—Content engines join the CDN and self organize to obtain preloaded content. Content engines will then communicate with their peers to monitor network conditions and determine the optimal paths for content transfer and redirection.
- Integrated video decoder—Content engines decode streaming video and audio for playback to a TV monitor. This facility is useful for conference rooms and kiosk applications.
- Freshness Controls—Content engines obey HTTP caching standards. System administrators may manually or automatically flush or preload content engine files. Content is delivered fresh.

Figure 12 Enterprise Content Engine Deployment



Intelligent Networking Services

The network must have content-level intelligence to achieve a rich media experience. Requests should be served based on content, location, and identity, not indiscriminately. A Content Delivery Network with intelligent network services accelerates content delivery at the same time it provides security, QoS, voice, VPN, Multicast, NBAR, and other services. Cisco Intelligent Network Services can bring content services to new markets. These intelligent services are the infrastructure for new content-rich applications.

Cisco Intelligent Network Services enhance traditional IP services at content level. Through Intelligent Network Classification, the network can distinguish traffic based on application content and context. Features such as Network Based Application Recognition (NBAR) on the Cisco 2600 Series, Cisco 3600 Series, Cisco 7200 Series routers, in addition to Context Based Access Control (CBAC) on Cisco PIX Firewalls can perform content aware inspection for determining quality of service (QoS) behavior or security requirements.

Conclusions

Content providers and Web viewers will migrate to the highest-performing environment. Intelligent content networking solutions offer performance and cost advantages that promise a Web environment that is fast, scalable, and cost-effective.

Cisco CDN product line is the industry's only complete content delivery solution, incorporating sophisticated, patented technology that offers network service providers the opportunity to capitalize on this explosive market opportunity.

Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

European Headquarters
Cisco Systems Europe
11, Rue Camille Desmoulins
92782 Issy Les Moulineaux
Cedex 9
France
www.cisco.com
Tel: 33 1 58 04 60 00
Fax: 33 1 58 04 61 00

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-7660
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems Australia, Pty., Ltd
Level 17, 99 Walker Street
North Sydney
NSW 2059 Australia
www.cisco.com
Tel: +61 2 8448 7100
Fax: +61 2 9957 4350

Cisco Systems has more than 190 offices in the following countries. Addresses, phone numbers, and fax numbers are listed on the Cisco.com Web site at www.cisco.com/go/offices.

Argentina • Australia • Austria • Belgium • Brazil • Canada • Chile • China • Colombia • Costa Rica • Croatia • Czech Republic • Denmark • Dubai, UAE
Finland • France • Germany • Greece • Hong Kong • Hungary • India • Indonesia • Ireland • Israel • Italy • Japan • Korea • Luxembourg • Malaysia
Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland • Portugal • Puerto Rico • Romania • Russia • Saudi Arabia • Singapore
Slovakia • Slovenia • South Africa • Spain • Sweden • Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela

8/00 LW