HP Internet Security and E-Commerce Solutions

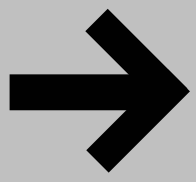
ESG11439SG0309

student

guide

HP Internet Security and E-Commerce Solutions

ESG11439SG0309



training

HP Internet Security and E-Commerce Solutions
Student Guide
September 2003

Contents

Overview

Module 1 — DISA Framework

Module 2 — Website Planning

Module 3 — Deploying Microsoft Commerce Server 2000

Module 4 — Deploying Microsoft Content Management Server

Module 5 — Managing Website Implementations

Module 6 — Implementing Load Balancing

Module 7 — Planning Internet Security

Module 8 — Intrusion Detection

Module 9 — Implementing Firewall Security

Module 10 — Implementing Virus Protection

Learning Check Answers

Glossary

Course Overview

Introduction

The HP Internet Security and E-Commerce Solutions course teaches you how to plan, implement, optimize, and secure an e-commerce website with e-commerce and security applications running on HP ProLiant servers. The recommendations in this course are based on proven best practices identified during extensive HP testing.

In this course you will build an HP an Internet security and e-commerce solution based on the HP Dynamic Internet Solutions Architecture (DISA).

Course Goal

You should learn to plan and design an e-commerce solution that provides high availability, scalability, and security. This will be achieved using:

- HP servers
- Software from HP partners
- Methodologies tested by HP

Note

This course does not address topics such as application design, development tools, or database design.

Course Objectives

In this course, you will implement a complete DISA architecture. At the end of this course you should be able to:

- Describe the components used in the HP DISA framework.
- List and explain considerations for planning a website
- Use applications to deploy Microsoft Commerce Server.
- Use applications to deploy Microsoft Content Management Server.
- Install, configure, and optimize load-balancing solutions.
- Describe the HP recommended best practices for securing a web server and managing the website content.
- Plan Internet security.
- Implement intrusion detection solutions using HP hardware and partner software.
- Plan and implement a firewall configuration using HP hardware and partner software
- Implement virus protection

Prerequisites

Before attending this course, you should:

- Understand the materials presented in the HP E-Commerce Fundamentals web-based training (WBT) and Internet Security WBT.
- Understand the setup, configuration, and general operations of Microsoft Windows 2000 Server and associated support and configuration applications.
- Understand general Windows 2000 Server system administration, including user-level security requirements, user account management, and application support and management.
- Be familiar with network administration and support, including hardware, software, setup, configuration, topologies, and security.
- Understand network technologies and protocols such as Transmission Control Protocol/Internet Protocol (TCP/IP), Domain Name System (DNS), Windows Internet Name Service (WINS), and Dynamic Host Configuration Protocol (DHCP).
- Be familiar with high-end server systems, options, and software, especially HP servers, HP hardware options, and HP software products.

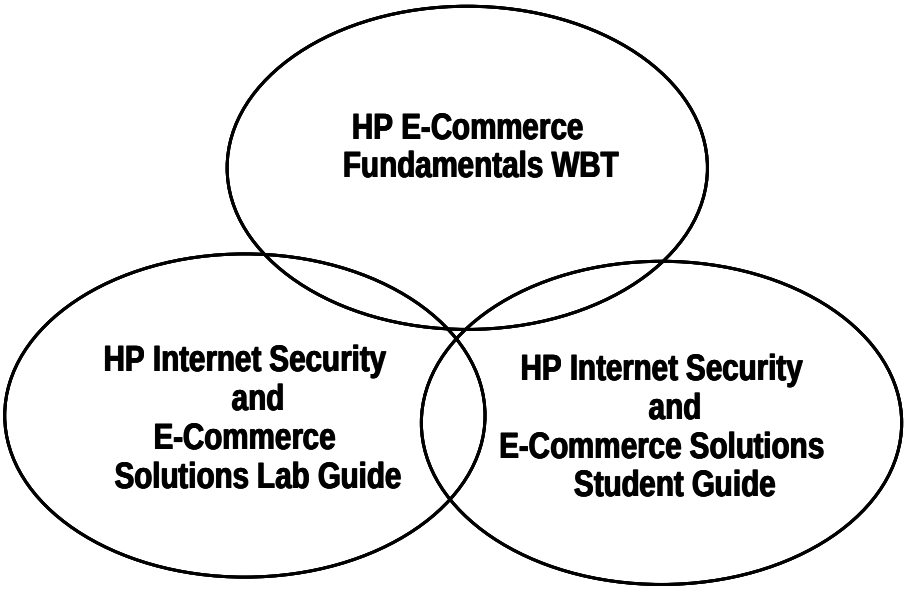
Note

These prerequisites are equivalent to the HP ASE certification (Microsoft Windows 2000 track).

It is preferred that you are a Microsoft Certified Systems Engineer (MCSE) and that you have attended the following HP courses or have equivalent industry experience and knowledge:

- HP Systems Technologies
- HP/Windows 2000 Integration and Performance

Course Overview



The HP Internet Security and E-Commerce Solutions course consists of the following key components:

- Web-based self-study training
- Student guide
- Lab guide

! Important

Certification exams based on this course will contain questions from all the preceding components.

HP Partnerships

Working closely with industry-leading independent software vendors, HP tests and integrates its Internet security and e-commerce solutions to provide optimal performance and exceptional reliability.

ActiveAnswers

ActiveAnswers is an online subscription service for HP partners and customers that delivers tools and information that enable you to implement standards-based computing solutions quickly, safely, and efficiently.

INTERNET

ActiveAnswers is located at:

http://www.hp.com/solutions/activeanswers

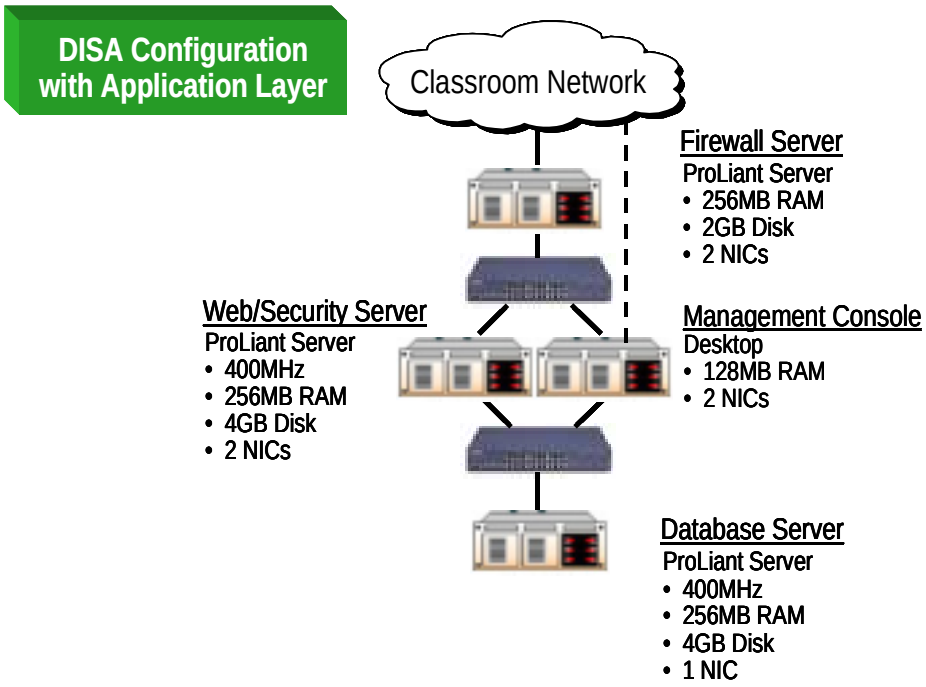
Hardware Configuration Options

Four types of servers exist within a workgroup:

- **Web/security servers** — HP ProLiant Servers that have these main functions:
  - Web servers
  - Intrusion detection server
  - Vulnerability assessment server
  - Virus detection server
- **Firewall Servers** — HP ProLiant servers that function as a firewall
- **Database servers** — HP ProLiant servers that store catalog data
- **Client/management console** — HP desktop that is used to configure and manage the web servers, firewall server, and database server

Three hardware specifications using the preceding types of servers can be used for the classroom configuration.

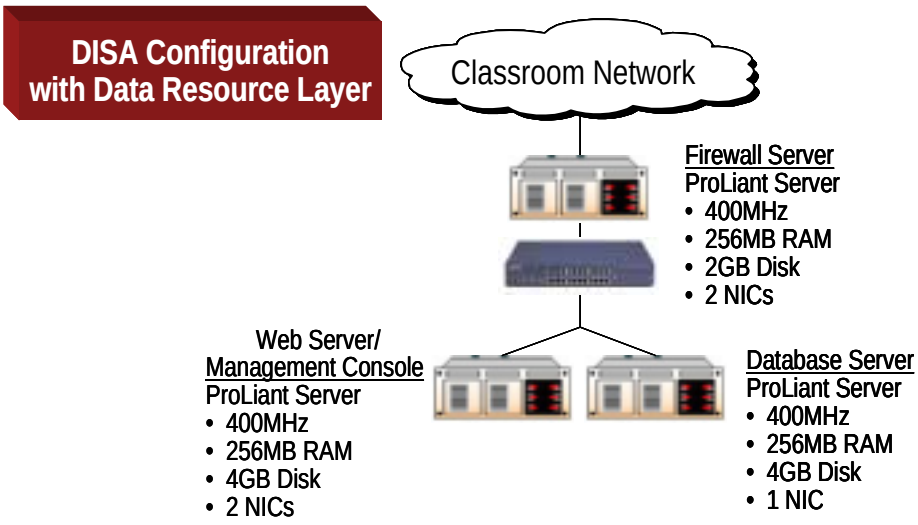
Hardware Configuration Option 1



The DISA configuration with the application layer separates the data resource content from the web server. The use of two web servers facilitates the implementation of load balancing within a single student group.

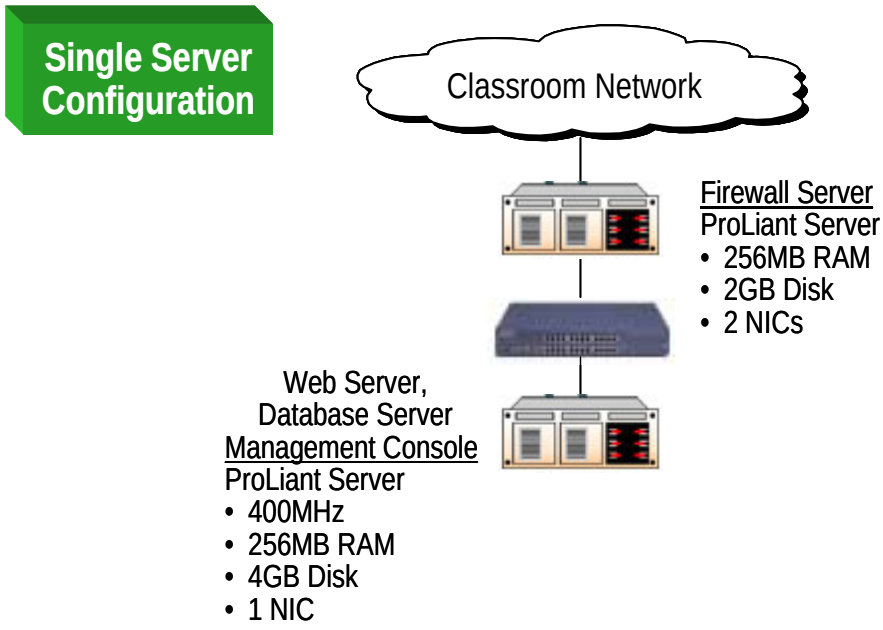
In addition, one of the web servers can be used as the management console that controls all servers from a central location. The other web server can also be used as a security server if the HP AXL300 Accelerator PCI Card or Secure Sockets Layer (SSL) is installed.

Hardware Configuration Option 2



The DISA configuration with the data resource layer separates the database content from the web server. This configuration requires that the web server be used as the management console and does not permit load balancing within a single student group.

Hardware Configuration Option 3



The single server configuration combines the database and application on one server. A second server is used as a firewall.

Software Configuration Overview

Server Type	Configuration
Web Server	Windows 2000 Advanced Server Service Pack 2 Windows 2000 hotfixes Internet Information Server (IIS) 5.0
Database Server	Windows 2000 Advanced Server Service Pack 2
Firewall	Windows 2000 Advanced Server Service Pack 2
Client/Management Console	Windows 2000 Professional Service Pack 2 Internet Explorer 5.5

The preceding table lists the software that is preinstalled on each server. You will install and configure additional software components during the labs.

Network Configuration Overview

The following table shows the basic network configuration and workgroup IP assignments for this course.

Group IP Configurations	
Group 1	
Firewall1 external IP address	192.168.1.10
Firewall1 internal IP address	10.1.1.11
Web1 IP address	10.1.1.12
Database1 IP address	10.1.1.13
Client1 external IP address	192.168.1.11
Client1 internal IP address	10.1.1.14
Root website	group1.com
Group 2	
Firewall2 external IP address	192.168.1.20
Firewall2 internal IP address	10.1.1.21
Web2 IP address	10.1.1.22
Database2 IP address	10.1.1.23
Client2 external IP address	192.168.1.21
Client2 internal IP address	10.1.1.24
Root website	Group2.com
Group 3	
Firewall3 external IP address	192.168.1.30
Firewall3 internal IP address	10.1.1.31
Web3 IP address	10.1.1.32
Database3 IP address	10.1.1.33
Client3 external IP address	192.168.1.31
Client3 internal IP address	10.1.1.34
Root website	Group3.com
Group 4	
Firewall4 external IP address	192.168.1.40
Firewall4 internal IP address	10.1.1.41
Web4 IP address	10.1.1.42
Database4 IP address	10.1.1.43
Client4 external IP address	192.168.1.41
Client4 internal IP address	10.1.1.44
Root website	Group4.com

Note

The IP addresses in the classroom have been selected for classroom convenience and software licensing requirements. They do not represent actual IP address implementations.

Credits and References

Hill, Brett, *Migrating from IIS 4.0 to IIS 5.0*. Windows 2000 Magazine. September 2001.

Compaq Computer Corporation. *The Compaq Dynamic Internet Solutions Architecture*. February 1999.

Compaq Computer Corporation. *Compaq TaskSmart C-Series Server and DISA Integration Guide*. January 2000.

Compaq Computer Corporation. *How Caching Accelerates Network Performance and Reduces Costs*. September 1998.

Compaq Computer Corporation. *Load-Balancing Considerations for Dynamic Internet Solutions Architecture Environments*. September 1998.

Compaq Computer Corporation. *Managing Multi-Tiered E-Business Applications*. February 2001.

Compaq Computer Corporation. *Microsoft Commerce Server 2000 Installation Checklist*. May 2001.

Compaq Computer Corporation. *Performance Characterization of Microsoft Commerce Server 2000 on Compaq ProLiant Servers for Storefronts*. October 2001.

Compaq Computer Corporation. *Scaling Up and Scaling Out with Compaq ProLiant Servers and Microsoft Windows*. August 2000.

Compaq Computer Corporation. *Solution Guide for Microsoft Commerce Server 2000 Storefront on Compaq ProLiant Servers*. May 2001.

Microsoft Corporation. *A Blueprint for Building Web Sites Using the Microsoft Windows DNA Platform*. Online. January 2000.

Microsoft Corporation. *Architectural Design: A Scalable, Highly Available Business Object Architecture*. Online. January 2000.

Microsoft Corporation. *Commerce Server 2000 Documentation*. Online. July 2001.

Microsoft Corporation. *Installing Commerce Server 2000*. Online. November 2001.

Microsoft Corporation. *Microsoft Commerce Server 2000: Planning for Reliability and High Availability*. Online. January 2001.

Objectives

After completing this module, you should be able to:

- Define the HP Dynamic Internet Solutions Architecture (DISA) framework.
- Describe the DISA components including the:
  - Core application stack
  - Global components
- Explain the steps to deploying a DISA implementation.
- Explain how Microsoft products can be integrated into DISA.

Introduction

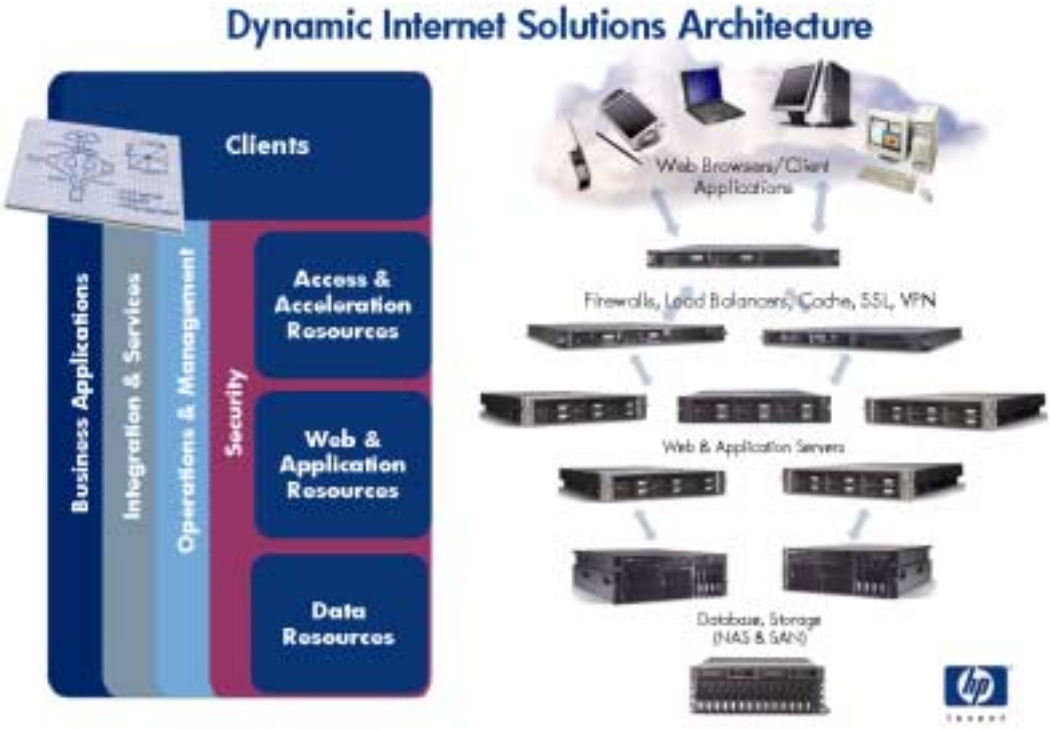
The explosion of innovative Internet technology is significantly influencing the way applications are written and deployed. The hundreds of thousands of Internet websites that were static “brochure-ware” are quickly becoming highly interactive Internet applications with transactional capabilities. Inside large corporations, developers are using web technology to integrate enterprise applications into large-scale intranets. Between corporations, business partners are building secure extranets to streamline their supply chains and improve communication.

As applications expand on the Internet, and on enterprise intranets and extranets, the functions they perform are becoming increasingly mission critical. The Internet order fulfillment applications must be scalable and available at any time of the day or night and on any day of the week. The electronic commerce (e-commerce) site must be robust and scalable to keep up with traffic growth and maintain a quality of service that does not compromise revenues. Security, availability, scalability, and manageability must be considered when designing the infrastructure of an e-commerce site.

In addition to the increasing scalability and availability features, leading Internet application architects are starting to use multitiered, distributed architectures. Rather than use a single server that performs all application functions, systems now involve many servers working together to deliver the application to users.

In response to the requirements of mission-critical Internet applications and the trend toward multitier distributed architectures, HP has developed an advanced architecture for building highly scalable, highly available Internet applications. HP DISA provides the basis for building scalable, highly available e-commerce solutions.

HP DISA Framework



The DISA framework is based on conventional scalability and availability principles and technology, but applies these technologies in new ways to fit the Internet application paradigm.

The HP DISA framework consists of a three core components and four global components. The core application stack components are access and acceleration resources, web and application servers, and data resources. The global components are security, operations and management, integration and services, and business applications. In most instances, the clients are web browsers; however, in today's business-to-business (B2B) environment, the client may be an application required at the Business Applications level.

This multitier architecture is a fundamental requirement for building a robust e-commerce solution. The DISA approach distributes loads across multiple servers, providing more scalability and availability than a single-server solution.

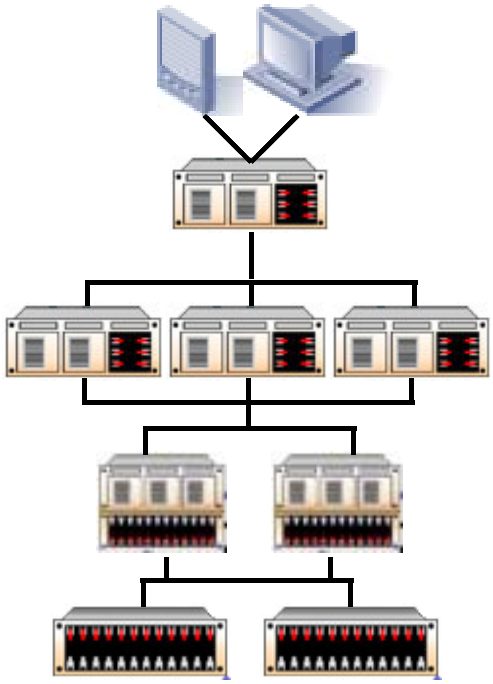
DISA Client Layer

Web Browsers,
Client Applications

Firewalls, Load Balancer,
Cache, SSL, VPN

Web &
Application Servers

Databases,
Storage
(NAS & SAN)



The client layer is where application data is consumed. Clients issue requests to a service name, which represents the application being delivered to the client. The user and the client software have no knowledge about the inner working of the system that delivers the service. The DISA framework applies to web browsers as well as other Internet applications that have non-browser clients, such as FTP and Internet email.

Core Application Stack

The core application stack of DISA contains the components that deliver the application functionality.

Access and Acceleration Resources Tier

The hardware and software at this tier provide the interface between the underlying site and the incoming network resources. For example, load balancers can provide one virtual host name to clients and distribute the load across multiple servers, firewalls can provide defense against intrusion, caching appliances can provide faster response times, and VPN (Virtual Private Network) appliances can provide secure connections between remote sites.

Load Balancing

There are multiple approaches to load balancing, from hardware to software solutions. One of the primary functions of the access and acceleration tier is to load balance the web and application servers.

Load balancing is done at either the transport layer or application layer of the Open Systems Interconnect (OSI) model. Transport layer load balancing uses base algorithms to determine which server will get the next request. In application layer load balancing, the request can be examined before deciding which server or set of servers should service the request.

Another approach is to load balance switches as a hardware solution. As the Internet becomes a more global tool, load balancing between sites becomes an issue.

Firewalls

The first line of defense in the security of a site is the firewall. Overall security of a DISA site includes multiple firewalls, intrusion detection, virus protection, and controlled access to the site using a VPN.

The firewall controls what can enter and leave the site. Firewalls may be deployed between each core component, with the firewall definitions being tighter at each level.

Caching

A caching appliance stores information for faster retrieval. Ideally, the caching appliance will service the request, eliminating the need for the web or application servers to handle the request. By handling a request for a specific file or information at the highest level, performance of the site can be improved.

Virtual Private Networks

VPNs are private, secure connections between remote sites on the Internet. A VPN may allow multiple incoming paths (if the user has the correct key) or it may be a single point-to-point connection. Traffic on a VPN is encrypted and secure. Many firewall applications now include VPN features and support.

Web and Application Servers

The servers at this tier provide web services functionality and perform the primary work of the applications for the site. For example, an Internet application is a basic web server that serves primarily static HTML pages interspersed with a few Common Gateway Interface (CGI) scripts that process HTML form input (for example, registering users on a mailing list).

More sophisticated Internet applications use advanced CGI scripts, Personal Home Page (PHP) scripts, Microsoft Active Server Pages (ASPs), or Java Servlets that run on the application server. The data for these applications can be replicated on each server or stored on separate servers in the data resource tier of the DISA framework.

Microsoft Commerce Server and Content Management Server provide functionality that retrieves and processes data from databases, integrates with other applications, and then returns formatted results to the web server.

Data Resources

The data resources tier servers contain the data for the applications in the web and application resources tier. Highly available, centralized file systems can store critical data on server clusters, Fibre Channel storage, Storage Area Networks (SANs), or network attached storage (NAS) solutions. The data may consist of a database, email, streaming video, or e-commerce information.

Most web servers cache HTML, images, and other data resources in memory to avoid the delays inherent in accessing local or network drives for each URL request. In a DISA framework using centralized data storage, the web servers fill their caches from network file shares rather than from local drives.

Like any centralized application resource, data resources must be highly available so they do not become a single point of failure. Cluster-enabled operating systems provide a relatively simple and cost-effective approach to file system and database availability.

Example

Although it is possible to install all Commerce Server components on one server, a multiserver configuration is required to separate the business logic and database tiers and to provide failure protection.

A four-server configuration has two servers at each tier: two identical web servers and two identical database servers. The web servers are load balanced and are running Commerce Server and the database servers are configured as clusters sharing a SQL Server database.

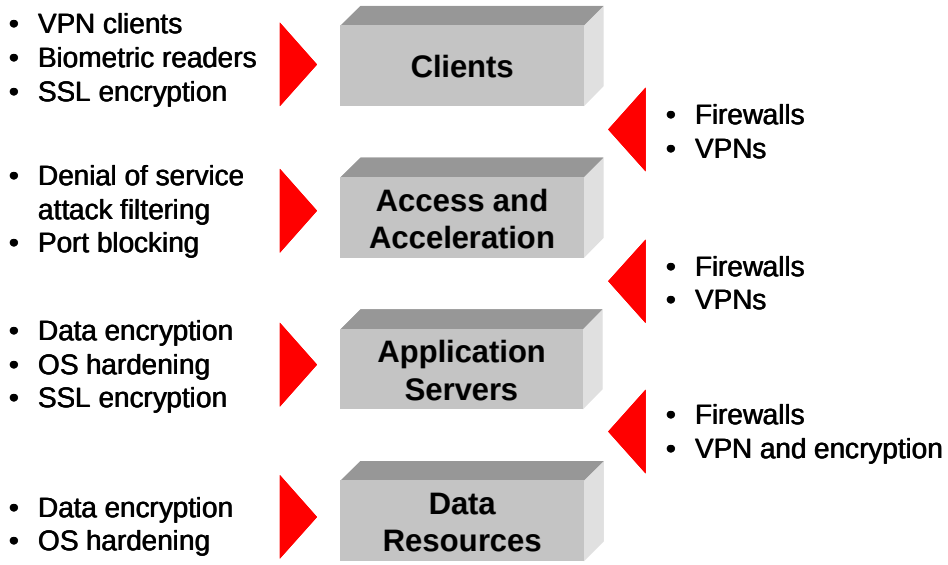
Global Components

In addition to the core application stack, the four global components of a DISA framework help complete the model. These components may include software and hardware solutions, but all are defined by a set of best practices for implementing each component.

Security

Mission-critical applications deployed on DISA systems typically require a robust approach to security. Ensuring privacy, thwarting denial-of-service attacks, and protecting data from theft are obvious requirements for any electronic business or extranet application that connects to the Internet. Many of these security concerns also apply to intranet applications.

Security can be implemented at each level of the DISA framework by using security technologies such as firewalls and various forms of encryption.



Some acceleration devices provide a base level of security protection for the rest of the DISA environment. Router-based load balancers block connections through all TCP/IP ports other than those the administrator chooses to open.

The way security technologies are implemented can significantly affect the scalability and availability of a DISA environment.

Example

If you place a firewall between the web servers and the data resources, a performance bottleneck can occur, depending on the security settings. Not all firewalls support proprietary database protocols, which can hinder the application servers from accessing the databases. Ensure that the firewall is properly configured to access all required resources.

Operations and Management

Operations and management best practices are used to maintain an efficient site by establishing and monitoring infrastructure, version control, and service level agreements. For example, HP Insight Manager 7, Integrated Lights Out (iLO) and HP Remote Insight Lights-Out Edition (RILOE) are recommended best practices for managing DISA configurations.

The following table summarizes some important elements of systems management in a DISA environment and suggests tools that can be helpful at each level.

Table 1

Function	Description	Examples
Hardware and Operating System Monitoring	Monitor status of load balancers, disks, processors, memory, network adapters, and Operating System	HP Insight Manager HP OpenView Computer Associates Unicenter TNG Tivoli TME
Remote Diagnostics and Administration	Diagnose server problems remotely Remotely control and administer a server	HP Remote Insight Board HP Integrated Remote Console HP Remote Insight Option HP Carbon Copy
Server Recovery	Automatically reboot failed servers	HP Automatic Server Recovery
Application Service Monitoring	Determine if an application service (web server) or data resource (database) is operating correctly	Freshwater Software SiteScope Bright Tiger Cluster Cats NetIQ AppManager
Performance Benchmarking	Test server and site performance	eTesting Labs webBench Keynote Systems Performance Appraisal Service

The HP RILOE II board and iLO allow complete remote management of a system over the network. In conjunction with the notification capabilities of HP Insight Manager, you can remotely determine the state of a system, upload software, install drivers, and reboot a server without touching the server. You can remotely access and manage a server from anywhere on the Internet if it is correctly configured to use a VPN.

DISA systems involve many different servers that work together to deliver a complete application. Server management has a significant impact on the availability, scalability, and total cost of ownership of DISA-based configurations. A robust approach to systems management is required to keep all the systems online and running at peak performance.

Integration and Services

HP engineering groups work with the HP services organization and service delivery partners to create solutions and define best practices for integrating applications and business-partner systems that use technologies such as SOAP, XML, and UDDI on HP hardware.

Discussions with the HP services organization and customers help define the types of solutions that are most beneficial. Working with the appropriate partners, HP creates an integrated solution of software and hardware following the DISA framework.

The creation of a complete solution that is highly available, scalable, manageable, and secure uses best practices that include:

- Scaling out the web and application resources tier by using multiple servers that are load balanced.
- Scaling up the data resources tier by using fully configured servers that may be clustered with external storage.
- Using component redundancy at potential single points of failure.
- Implementing necessary firewalls at each component level.

Business Applications

Solution performance is contingent on the application. Performance is impeded if an application is not written to take advantage of the DISA framework. Best practices for this area provide guidelines for optimizing the performance of applications in DISA.

Example

Maintain user-state information in an easily accessible location such as client cookies, database tables, and URLs instead of on a single application server. As the site grows, application design will have an increasing impact on performance.

Deploying a DISA Implementation

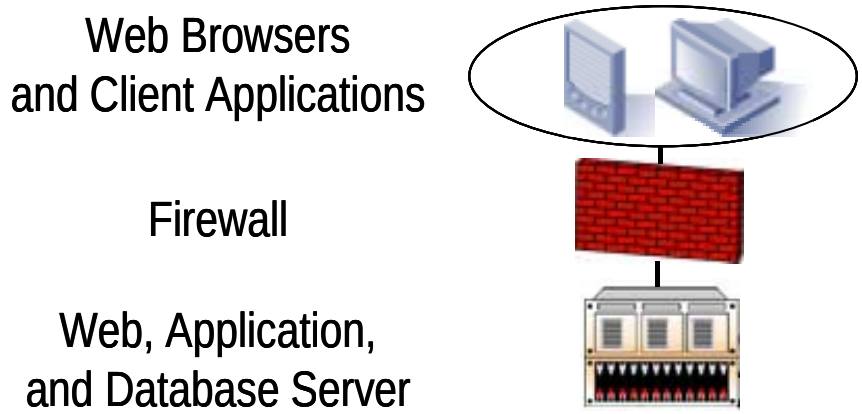
A DISA implementation can be deployed on a modular basis, allowing components to be added or upgraded without drastically altering the complete solution.

The deployment steps are:

1. Implement a single server.
2. Move the content to a dedicated server.
3. Add load balancing.
4. Add web and application servers.
5. Add redundancy and capacity.
6. Eliminate single points of failure.

Following is an overview of each step of the DISA implementation process.

Implementing a Single Server

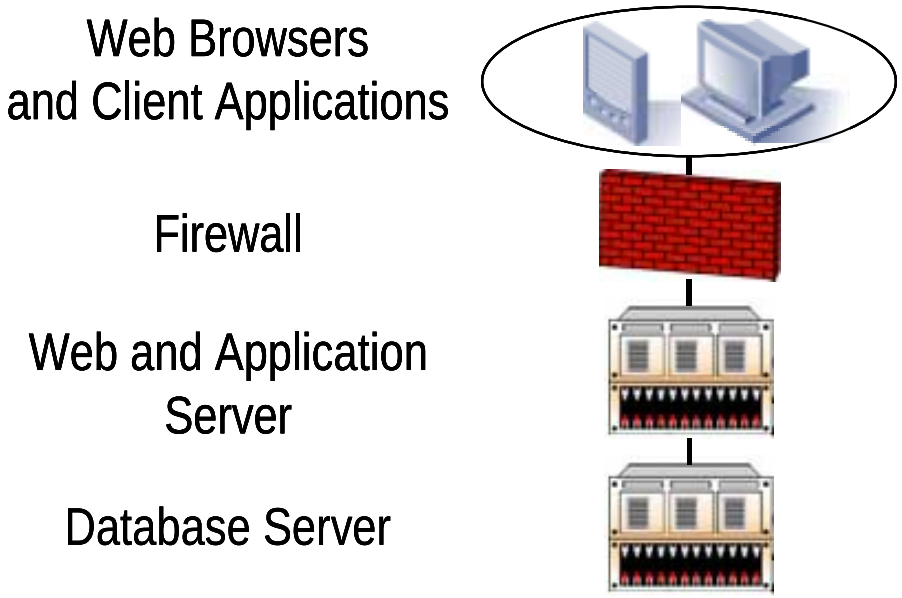


With a stand-alone server configuration, the application server and data resources are often combined on a single server. Such servers may simultaneously run a web server such as IIS, an application such as Commerce Server 2000, and a database such as Microsoft SQL Server.

You can add processors or memory to improve the performance of a stand-alone server. However, many applications scale with decreasing returns as processors are added. There is always an absolute limit to how many memory modules and processors a single server can handle. Some applications may never tax an amply configured stand-alone server, but even moderately complex CGI or ASP applications can quickly reach the processing power barrier.

Availability features such as RAID disk arrays, error checking and correcting (ECC) memory, and redundant power supplies, fans, and NICs can help reduce downtime of a stand-alone server. These features protect against many of the most common hardware failures. Features such as the HP Automatic Server Recovery (ASR) can detect operating system failures and automatically reboot the system.

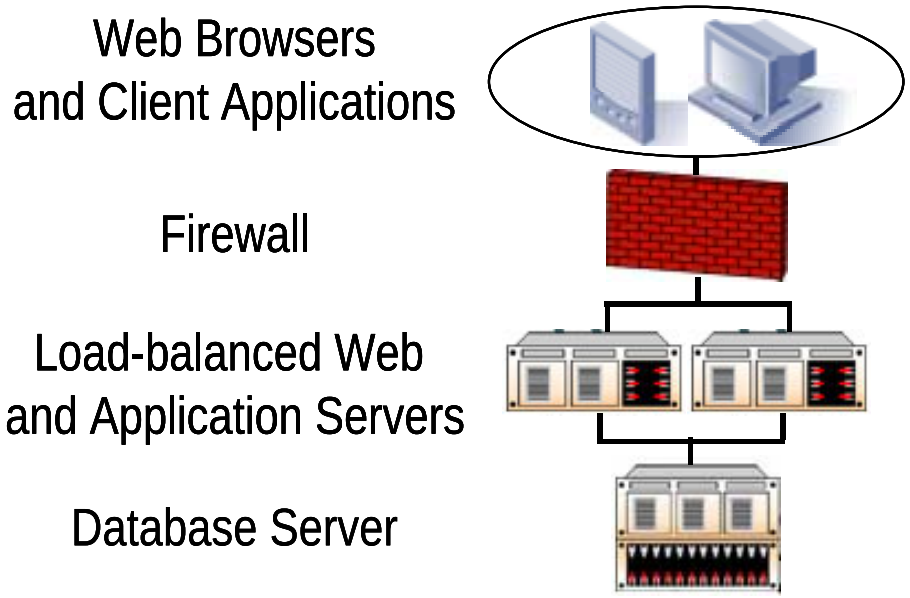
Moving Content to Dedicated Server



File system-based resources (such as HTML and images) can be stored either on the local drives of application servers or on the database server. The main consideration in determining where to store the web server files is network traffic and performance.

Storing all data on a dedicated database server eliminates the problem of keeping replicated data synchronized. This is a significant challenge with most successful websites because the data resources are large and changing. However, if network performance between the back-end systems and the web and application servers becomes an issue, replication of the web data to the web servers may become a necessity.

Adding Load Balancing



The third step in deploying a DISA implementation involves adding the load-balancing component. Load balancing distributes computing tasks and client requests across multiple application and web servers. The result is a system that can maintain acceptable response times for customers.

Although the load-balancing component can be configured to face the Internet, this configuration is not as secure as other configurations. The typical configuration is to place a router, proxy, or firewall between the Internet and the load balancer.

If a web application requires clients to establish and maintain a session on a single web server, load balancing must be configured to enable affinity.

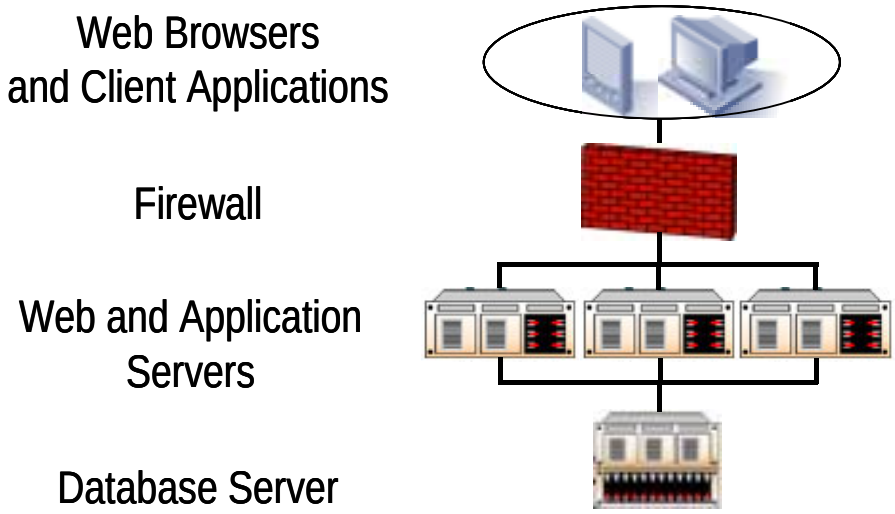
Example

If an AOL user who has placed multiple items into a shopping cart is then disconnected, because of the AOL proxy configuration the user might return to the shopping site using a different IP address than was used in the original connection.

If the load balancer can recognize a previous session based on multiple scenarios, such as TCP/UDP, SSL ID, and cookies, the shopper will be placed back on the same application server as before.

If the load balancer does not support multiple ways to identify a previous session, the user's connection might be assigned to the next available server and the previous shopping cart information will be lost. Thus, the need to maintain a persistent session is not met by the appliance load balancer.

Adding Web and Application Servers



An application server layer adds modularity to the solution. If load balancing is working properly, you can modify the application or web server layer without adversely affecting the site or taking the site offline.

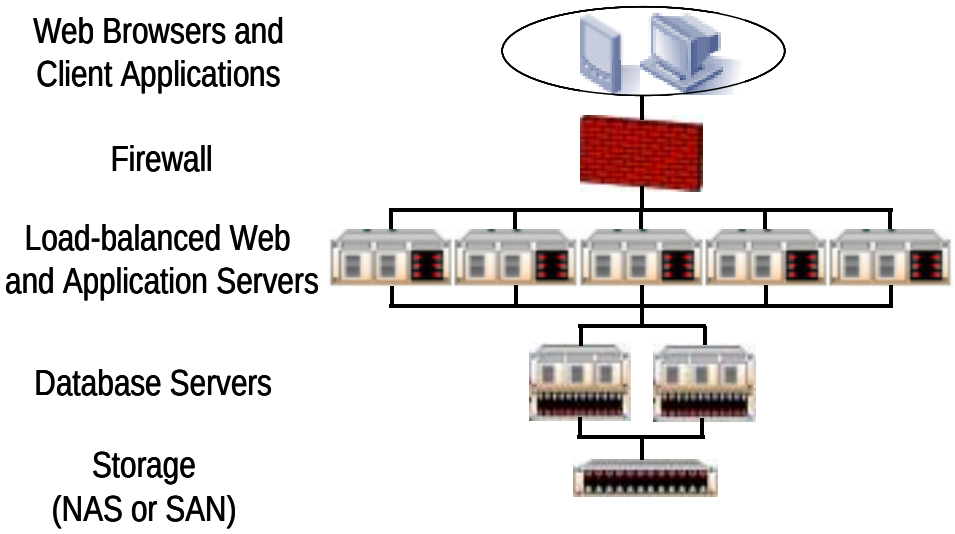
When combined with intelligent load-balancing technology, multiple application servers are more scalable and available than a single server with many processors and extensive memory. In cases where availability is critical, multiple servers in a load-balanced configuration allow a single server to be serviced offline while keeping the other servers online at all times.

INTERNET For more information download Scaling Up and Scaling Out with ProLiant Servers and Microsoft Windows from:

http://www.hp.com/solutions/disa

Application servers may have one or more processors depending on the characteristics of the particular application. By using appropriate load-balancing technology, it may be possible to combine servers with different processing power, I/O characteristics, or even different operating systems in an application server array.

Adding Redundancy and Capacity

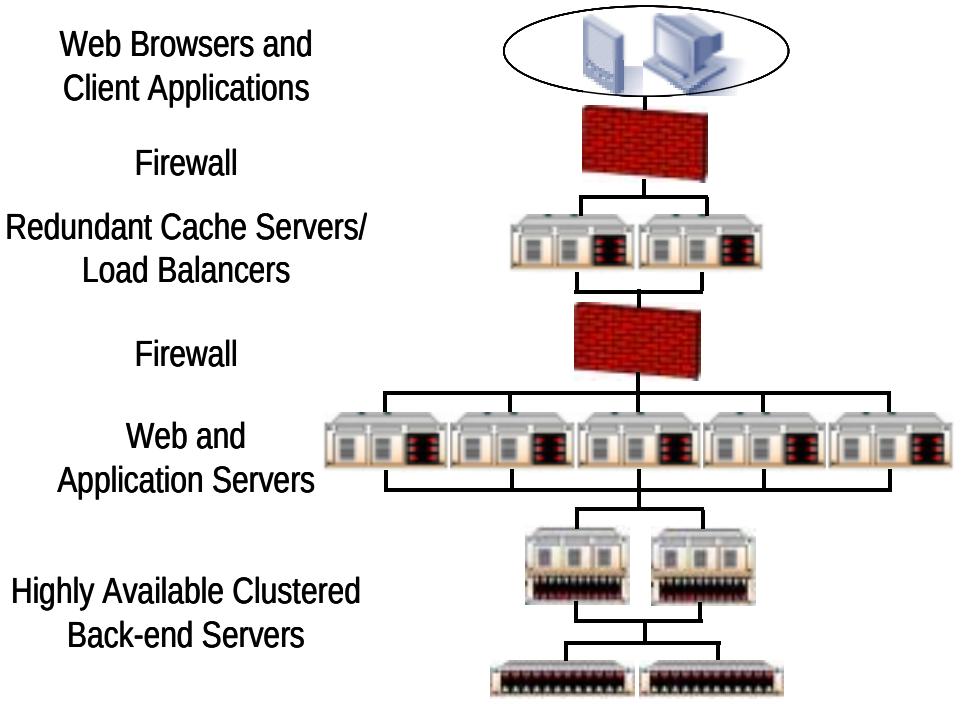


Centralizing data resources on fully configured, highly available servers rather than replicating files on the application servers may be the most efficient approach to ensuring redundancy and capacity. The impact of the data traffic affects the overall performance of the site. If appropriate failover techniques are applied, centralized data resource servers can simplify the task of managing rapidly changing application resources. Centralizing application resources also simplifies capacity expansion and failure recovery.

To minimize downtime and improve performance, data resource servers should be fully configured with multiple processors, have plenty of RAM, and use redundant components (NICs, power supplies, and fans).

Clustering technology can significantly improve the availability of centralized resources. However, depending on business requirements and application support, centralization may not always be appropriate. Performance gains of replicated data may offset the support and maintenance required.

Eliminating Single Points of Failure

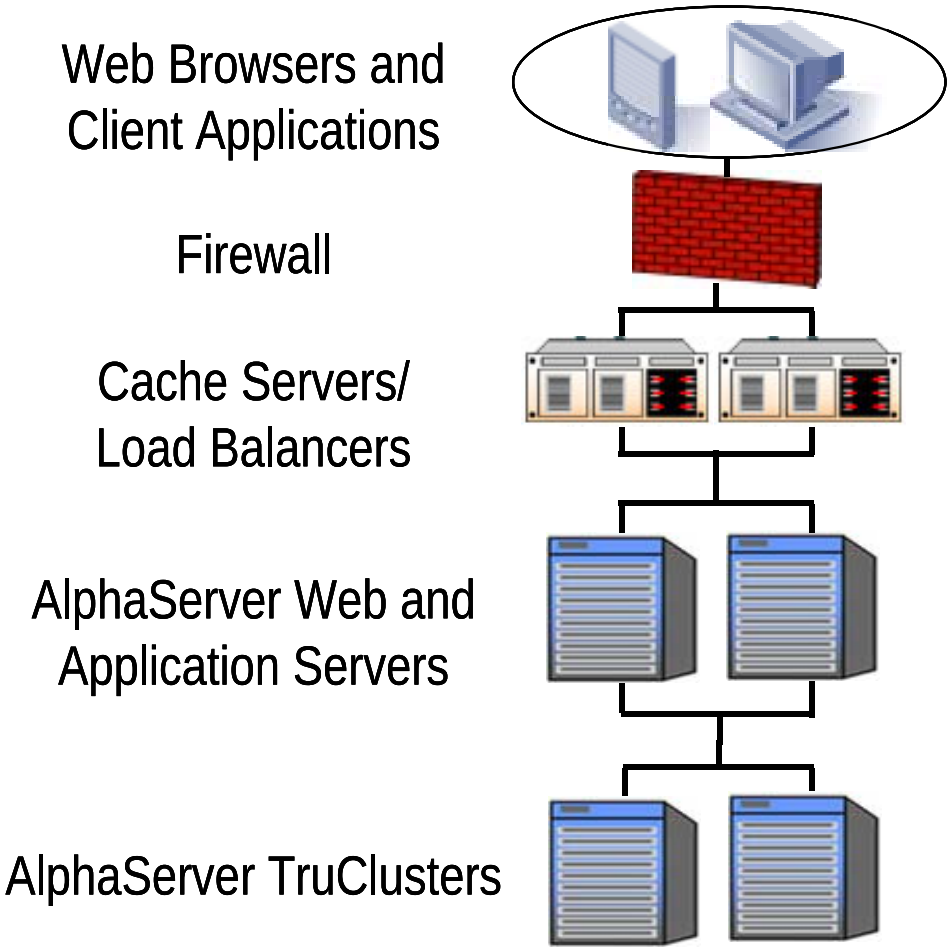


The correct degree of redundancy depends on the underlying business requirement and the available budget.

- Use failover configurations for load balancers, firewalls, and networking devices.
- Use intelligent load balancing technology provides failover for redundant application servers.
- Use clustering databases and network storage to enhance uptime for data resources.
- Use redundant NICs, power supplies, and fans to enhance uptime on data resource servers.
- Use multi-homing and an uninterruptible power supplies (UPSs) to provide redundant network connectivity.

Eliminating single points of failure is the final step to completing the architecture. However, tuning the solution for continuous availability and increasing capacity is an ongoing process. The DISA framework makes the process easier, allowing scalability without altering the entire solution.

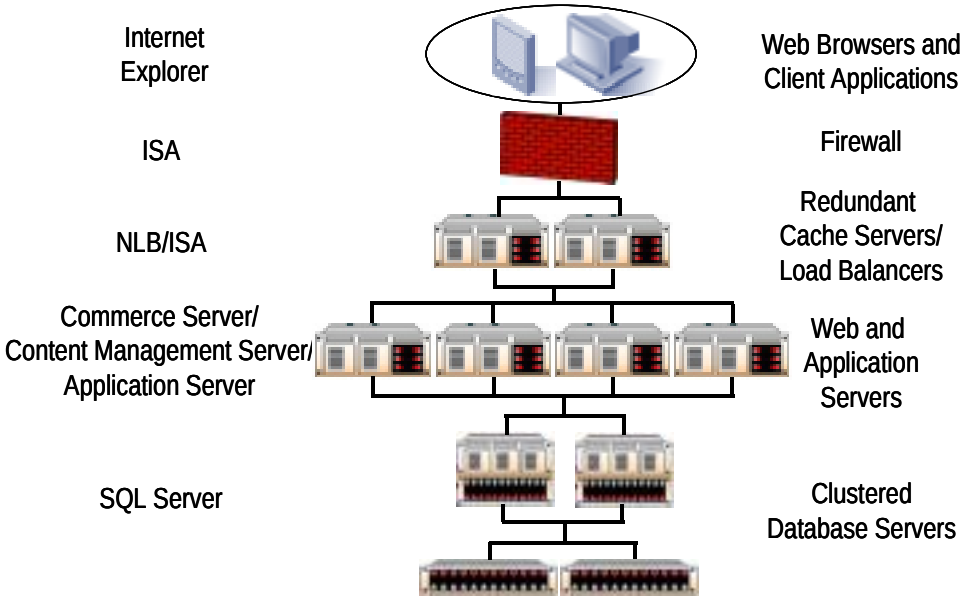
Alternate DISA Configuration



Building the DISA framework is not limited to a particular hardware or software solution. As indicated in this diagram, many alternate components can be integrated into the architecture.

The components of the DISA framework may vary in type and configuration. Remember that the functional principles of DISA are to distribute the work load to the application server, create a single back-end data resource for data replication purposes, and create availability, scalability, security, and manageability.

Integrating Microsoft Products in DISA



The following software components can be integrated into the DISA framework to build an e-commerce solution:

- **Internet Explorer** — Provides the web content interface to the clients.
- **Internet Security and Acceleration Server (ISA)** — Provides the security component.
- **Windows 2000 Network Load Balancing (NLB) service** — Enhances the availability and scalability of web server applications.
- **Commerce Server and Content Management Server** — Provide the application layer where the majority of processing in a solution takes place. Commerce Server also provides integration to the back-end data resource.
- **Application Center** — Provides a console to deploy and manage web applications and load balancing.
- **SQL Server** — Provides the back-end or data resource component.

Learning Check

1. Describe the basic principles of the DISA framework.

   .....

   .....

   .....

   .....

   .....

   .....

2. List the components of the DISA.

   .....

   .....

   .....

   .....

3. What are the steps for deploying a DISA implementation on a modular basis?

   .....

   .....

4. What Microsoft components provide services at the web server layer of the DISA framework?

   .....

   .....

Website Planning

Module 2

Objective

After completing this module, you should be able to describe the steps in planning a website or e-commerce solution that include:

- Establishing an online identity
- Planning the interface
- Selecting site building tools
- Assessing hardware requirements
- Addressing potential performance bottlenecks

Planning a Website

With the rapid growth of the Internet, a web presence is important for a business. However, before diving into developing a website, you must be clear about the goals of the site. What is the intended purpose or need? Are you planning to implement an intranet for employees, an information/customer service Internet site, an e-commerce storefront for consumers, or a business-to-business (B2B) site for suppliers and business customers?

You must make some planning decisions before developing and deploying your e-commerce storefront. The types of sites you plan to implement dictate the amount of hardware required, software applications, and necessary modifications to your security policy.

Planning your e-commerce implementation is the first and perhaps most crucial phase in the Planning-Testing-Deployment-Management cycle. If your site is carefully planned, taking security into account, you will be able to build an e-commerce solution that enables customers and suppliers to make secure transactions without unexpected interruptions.

Whether the e-commerce implementation is business-to-consumer (B2C) or B2B, several factors should be addressed as part of the planning process:

- How many web servers are needed?
- Should one large multiprocessor server or several smaller load-balanced servers be used?
- What type of servers should be used?
- How much memory should be in the servers?
- What are the expected performance bottlenecks and how are they resolved?

Before implementing a website you must:

1. Establish an online identity.
2. Plan the interface of the site.
3. Choose the necessary site building tools.
4. Assess the hardware requirements.
5. Address potential performance bottlenecks.

Establish an Online Identity

The name of your company website defines who you are and what you do. With this in mind, choose your domain name carefully. As in product branding, your domain name communicates and reinforces the name of your business. When selecting a domain name:

- **Make it easy to remember** — Avoid complicated spellings or word combinations.
- **Keep it simple** — Make the site name the same as the name of your business. The key is for customers to be able to remember it after typing it into their browsers or seeing it once.
- **Register it quickly** — Register the website as soon as possible to ensure you get the best domain name for your business. It is also best to register similar names as well as misspelled variations so that customers will be directed to your site rather than receive an error message.

Planning the Interface

When you are building a site from the ground up, you should use a technique referred to as “storyboarding” before you begin coding your site. Storyboarding is the process of creating an overall plan showing structure, sequence, connections of each logical step and the details surrounding web pages or nodes.

Whether it is an Intranet, e-commerce storefront, or a B2B solution, you must determine the goals and objectives of the site. In either case, plan the structure of the site so that visitors can easily:

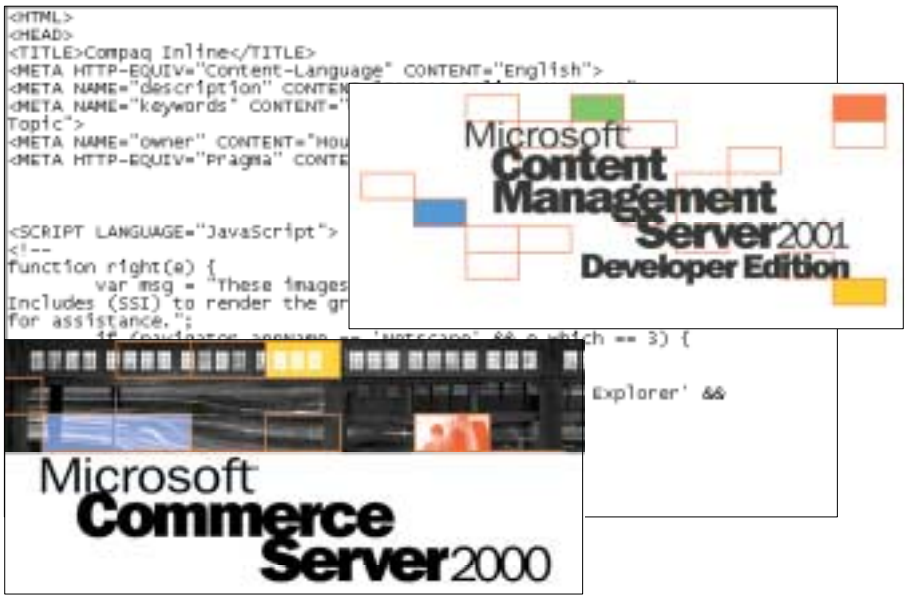
- Learn what they need to know.
- Locate products and information.
- Make purchases.

Typically, intranets are designed without regard to bandwidth and response time unless some visitors will be accessing the site over a slow WAN connection. If this is the case, you may want to avoid the use of unnecessary graphics, which can increase the time it takes to download a web page.

E-commerce and B2C initiatives are becoming enterprise-wide priorities and assets as the Internet sales and communications channel becomes mainstream. More than just websites, these customer-facing solutions provide a means of constructing and strengthening customer relationships by continually building trust based on multiple transactions and repeated provisions of value and reliability.

With restricted IT budgets, businesses are focusing on projects that improve customer relationships, increase customer retention, and decrease cost-per-sale while delivering information and business functions to a global audience. A B2B solution produces dynamic websites that provide information and services to a global audience with or without a transactional e-commerce component.

Selecting Site Building Tools



Some businesses use professional designers to create their websites. However, if your budget is limited or you have the capability to construct your website there are many ways to create a polished, professional site.

Building a traditional “page-at-a-time” website requires less planning and facilitates faster deployment. However, also requires frequent updates and has high maintenance costs.

When you build a website using Microsoft Content Management Server (MSCMS), the planning phase may take a little longer, but the development and maintenance phases are significantly reduced. MSCMS is a template-based system that separates the development of the structure of the site and the development of the content. One team of web developers, designers, and project managers can build the site framework and then hand it over to nontechnical content experts who populate the site with content. This allows these content experts to be self-sufficient and decreases the time required for site maintenance.

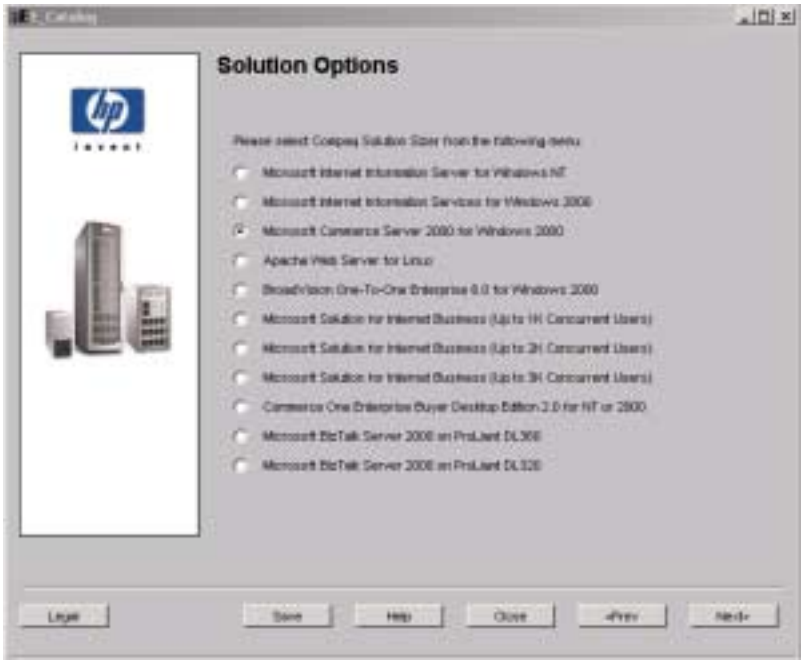
Microsoft Commerce Server is a comprehensive, integrated commerce solution that provides ready-to-use features and tools to quickly develop, deploy, and manage commerce applications for the Web. You can use Commerce Server to build B2C and B2B site applications.

MSCMS and Commerce Server require access to a SQL Server database. Databases are critical for storing website content, customer profiles, and product information.

Assessing Hardware Requirements

Review the hardware requirements for the development tools you selected. Keep in mind that the hardware configuration must also be able to accommodate website activity and maintenance.

Most small and medium e-commerce business owners do not have the technical background or the staff to support electronic commerce. An Internet Service Provider (ISP) or web hosting company can be used to connect your site to the Internet using one of their web servers.



If you prefer the complete control of purchasing, setting up, and managing your web server hardware and software, the HP E-Commerce Solution Sizer can assist you in sizing a solution environment for ProLiant servers.

INTERNET ActiveAnswers Solution Sizers are available at:
http://www.hp.com/solutions/activeanswers

The E-Commerce Solution Sizer provides server configuration guidelines for deploying Commerce Server on a variety of ProLiant platforms and StorageWorks solutions. When implementing a Commerce Server solution, you must determine if you will use a single server or DISA implementation.

With a single-server solution, all content resides on one server. This is an excellent solution for small to medium businesses or for internal use as intranet web servers. The single-server design is a vertical scaling architecture in which additional components such as memory and processors are added for scalability.

When availability, scalability, and fault tolerance are of utmost importance, a DISA solution that implements multiple application and data servers may be more suitable. With a DISA implementation, scalability to handle additional traffic is as simple as bringing another server online.

Addressing Potential Performance Bottlenecks

A bottleneck is caused by hardware or software that is operating at maximum capacity. As the load approaches maximum capacity, the bottleneck begins to restrict the flow of work through the system.

Performance tools can help you determine what hardware or software has reached a bottleneck. You can then upgrade the hardware, change the configuration, or tune the software to improve overall performance.

Although you cannot predict performance before you deploy your site, you can make sure you have a robust site architecture. This includes the installation of a Windows 2000 server that has the latest service pack updates. If you are planning a DISA implementation, Windows 2000 Advanced Server is required.

You should be able to gauge the amount of potential traffic on the website based on the site goals. If the website will be an intranet, you can project maximum traffic based on the number of individuals who have access to the site. Otherwise, an extranet or e-commerce site is purely customer driven. Site traffic estimates can be determined by market forecasts.

Learning Check

1. You are planning an intranet website for an advertising firm. The remote locations will access the intranet over a slow WAN connection. The web content will change often include many images. What are the considerations in planning the interface?

   .....

   .....

   .....

   .....

   .....

2. Discuss the implications of manually building web pages for an intranet.

   .....

   .....

   .....

   .....

   .....

Deploying Microsoft Commerce Server 2000

Module 3

Objectives

After completing this module, you should be able to:

- Describe the Microsoft Commerce Server 2000 architecture.
- Describe how the e-commerce business process integrates with Commerce Server.
- Describe the features and tools provided by Commerce Server.
- List and describe the administration tools used to manage Commerce Server.
- Describe the reports provided by Commerce Server to manage an e-commerce website.
- Describe the results of HP testing of Commerce Server for horizontal and vertical configurations.

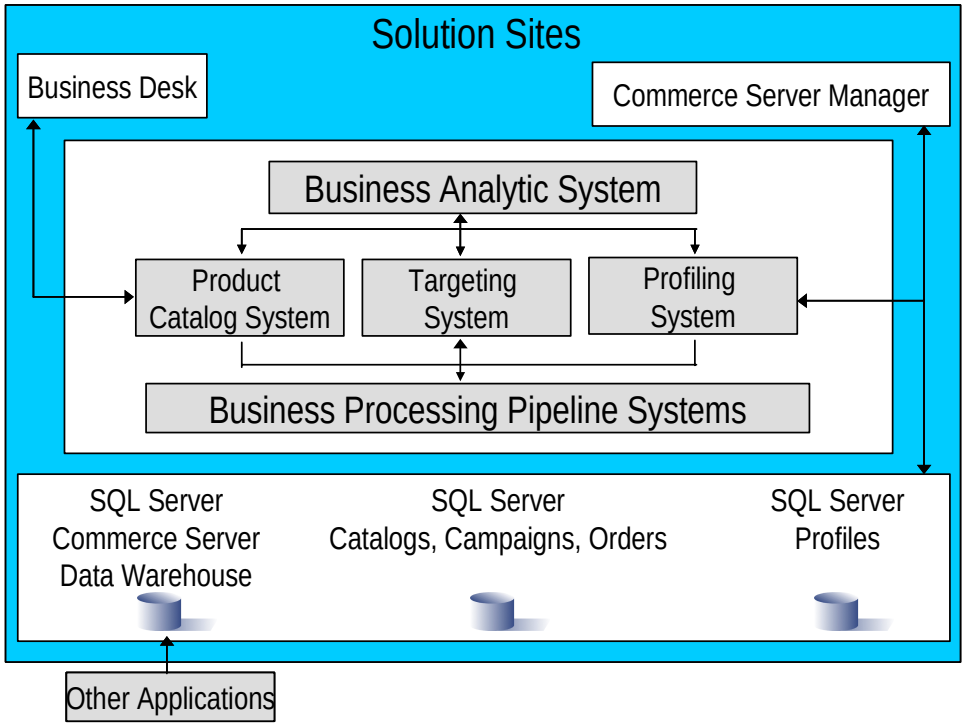
Overview

Microsoft Commerce Server 2000 provides a rich and extensible foundation for distributing tasks across the web servers. A website application can be developed to retrieve and process data from several different databases, then return formatted results to the web server.

Commerce Server is completely compatible with the Dynamic Internet Solutions Architecture (DISA). However, Commerce Server solutions alone do not ensure availability of the web servers. DISA covers availability and scalability gaps that Commerce Server architecture does not address.

Commerce Server Architecture

The Commerce Server architecture is based on the Component Object Model (COM). Commerce Server is designed for building tailored, scalable e-commerce solutions that optimize the user experience and provide business managers with real-time analytics and control of their online businesses.



The Business Desk component works with the Commerce Server systems to, for example, update catalogs, target content to users, profile users and organizations, and analyze site usage and productivity. The Commerce Server Manager component is used to configure system resources and manage applications, databases, and web servers.

The Business Analytic System, Business Processing Pipeline Systems, Business Desk, Commerce Server Manager, and the Commerce Server databases are customized and packaged in the Solution Sites, which can be used as a starting point for building a custom site.

E-Commerce Business Process

The fundamental reason to create an e-commerce site is to process orders. Commerce Server consists of series of distinct processes needed to process a customer's order, including:

- Computing tax
- Shipping an order
- Reconciling an order with an inventory database
- Integrating with existing business systems

These processes provide the basis for pipelines.

A pipeline is a software framework that defines and links stages of a business process, running them in sequence to complete a specific task. Commerce Server includes the following pipelines that perform the processing required by an online store.

Type	Description
Content Selection Pipeline	Determines which advertisements a user should see or which discounts a user should receive
Event Processing Pipeline	Records events, such as a user clicking an ad, that can be used for reporting purposes
Direct Mailer Pipeline	Constructs and delivers personalized email messages or static content
Order Processing Pipeline	Retrieves the product description from the catalog, retrieves the shipping address for the user, computes the amount of tax, and computes the total price

A site developer can tailor these pipelines to meet the specific needs for the site.

Content Selection

When you visit a website, the web server may create information about your visit or display personalized content. Commerce Server provides the tools necessary to target and track visitors.

Commerce Server provides two main components to help you engage and interact with customers:

- Profiling System
- Targeting System

Profiling System

Companies conduct market research to learn about customers. However, this can be a very costly, time-consuming process. When a customer makes an online purchase, you can collect information on that customer by building a customer profile. This can done by:

- Analyzing purchases.
- Logging which site was visited before yours.
- Gathering personal information through fill-in forms.

Targeting and Personalization

Targeting and personalization are the processes of delivering personalized content to users that have a profile. You can target content to anonymous users, registered users, companies, or contexts.

Example

You run a sports website, and you want to deliver tennis-related information only to tennis players or users who are interested in tennis. The information might include announcements about upcoming tennis tournaments and advertisements and promotions for tennis products.

To accomplish this task, use the Campaigns modules in Commerce Server Business Desk to perform expression-based targeting (explicit targeting). To perform expression-based targeting, you must know the profile properties of users to whom you intend to deliver content, offer promotions, or display discounts.

Prediction, or implicit targeting, is used when you do not have profile information on the user, or you do not have specific content to deliver for a particular profile. Use the Commerce Server Predictor resource to extrapolate this information from existing user data so that you can deliver content of interest to each user.

Expression-based targeting and prediction improve your ability to deliver content to specific users. Users are more likely to return to your site when they can expect to receive a personalized experience.

Event Processing

Commerce Server helps you to create highly dynamic, personalized commerce sites in which you can optimize the customer experience, encourage repeat business, and successfully compete in the e-commerce marketplace.

Cookies

Information created about your visit to a website is known as a cookie. A cookie contains information such as user ID, the date of your last visit to the site, and a time window that indicates how long the cookie is valid. Commerce Server uses cookies to identify and authenticate users and to associate user IDs with the profile information it collects about them.

Commerce Server creates two types of cookies:

- Persistent cookies are used to track anonymous users who visit a website. If an anonymous user allows cookies, the cookie is stored on the computer of the user. A persistent cookie is typically valid for several years.
- Nonpersistent cookies (session cookies) are used to track authenticated users who visit a website. When a user logs off a site, the session ends, authentication expires, and cookies are deleted.

Both persistent cookies and nonpersistent cookies can be encrypted.

Direct Mail

Websites can send bulk email messages to millions of users that include attachments, such as pictures, discounts, or URLs. The Direct Mailer service is a powerful engine that can send a large number of emails very quickly.

Example

You want to create a direct mail campaign item to distribute email message ads to a targeted group of users. The objective is to achieve a high response rate from the mailings.

To set up a direct mail campaign, use the Campaign Manager module in Business Desk.

Opt Out Lists

Some recipients do not welcome email advertisements and promotions. To accommodate these users, use the Direct Mailer opt-out feature. Individual recipients can request to be removed from a direct mail campaign list. Using an opt-out ASP page included in the Solution Sites, you can filter a mailing list against an opt-out list.

Order Processing

When you browse an online store and select which items you want to buy, you are accessing a catalog system. Commerce Server uses catalogs to organize and manage product data for display on a website and to target products to users. Catalogs are stored in a database.

To keep track of your selected items, the website uses some form of a virtual shopping cart system to compile your order. You can add or remove items from the shopping cart before finally clicking a checkout button to confirm your completed order. These steps use the Order Processing Pipeline.

Discounts

Customers can receive discounts on a particular item in the catalog, against an entire order, or on specific items an order.

Example

A customer visits the HP website and selects a notebook, leather carrying case, and an additional AC adapter. Free shipping is offered for all notebook purchases. In addition, customers receive a 50% discount on a 25-pack spindle of blank CDs and a CD holder when they buy a CDRW drive.

In Commerce Server the free shipping offer is an order-level discount because it applies to all products in a shopping basket. Free shipping is the only order-level discount delivered with Commerce Server Solution Sites.

Other types of order-level discounts, such as a 10% discount for preferred users, require customized programming. The 50% discount offer is considered a leverage discount in Commerce Server. These discounts are based on the purchase of other products.

The difference between percentage-off discounts and dollar-off discounts is that percentage-off discounts reduce the price of products by a percentage, rather than by a fixed dollar amount.

Checkout

After you have decided to make an online purchase, you enter profile information, which are personal details such as your name, address, credit card number, and other information, such as your email address. You may be asked to create a user ID and password, which will also be stored in your profile.

Credit Card Verification

Online retailers typically send a request to a financial institution to verify credit card account details, debit the buyer's account, and credit the retailer's account. If the transaction is approved, Commerce Server can provide the information for other business processes such as inventory, accounting, and fulfillment. When these tasks are complete, the order is processed and the items delivered.

The entire transaction takes place in real time, which means customers receive confirmation that an order has been processed in a browser page or with an email confirming the order and giving an expected delivery date.

Commerce Server Features

Commerce Server provides the critical e-commerce infrastructure needed to build an effective online business. Commerce Server provides the application framework for a secure, reliable, and comprehensive Internet commerce platform.

Commerce Server lets businesses:

- **Build** — Implement a commerce site that is easy for customers to navigate, find products, and take advantage of special discounts.
- **Engage** — Create and manage profile information, which can be used for targeted merchandising, personalized discounts, advertising, and adding predictive capabilities to the website.
- **Analyze** — Collect site usage data to determine site effectiveness, identify trends, distinguish specific groups of customers to target merchandise or content.

INTERNET

This section is intended to provide an overview of the features and capabilities of Commerce Server. For a complete explanation of Commerce Server, download the Commerce Server 2000 product documentation from:

http://www.microsoft.com/commerceserver/techinfo/productdoc/

The following table explains the features of Commerce Server in relation online commerce.

Commerce Server Feature	Activity
Product Catalog System	▪ Create and update catalogs of products to sell. ▪ Import catalogs of products and export catalogs in XML format. ▪ Browse by product category, name, or property value. ▪ Build custom catalogs with special pricing for groups of users. ▪ Apply discounts to products. ▪ Identify which products are selling. ▪ Exchange catalog data with vendors using Microsoft BizTalk Server 2000.
Profiling System	▪ Manage millions of users and organizations. ▪ Build profiles for users and organizations. ▪ Enable users to manage their own profile information and to research the status of their order. ▪ Analyze profile information to determine who is visiting your site. ▪ Create target expressions to match content to users. ▪ Create advertising, discount, and direct mail campaigns to target specific user groups.
Business Analytics System	▪ Import and manage site usage data to identify trends and new customer segments. ▪ Analyze site effectiveness. ▪ Target merchandise or content to specific groups of users as they navigate your site. ▪ Review collected data to identify hidden trends and new customer segments.
Targeting System	▪ Personalize the buying experience with targeted merchandising. ▪ Deliver content based on user context. ▪ Create, analyze, and manage personalized and targeted discounts, and direct marketing and advertising campaigns. ▪ Add predictive capabilities to your Website. ▪ Target ads or discounts to users of a specific profile. ▪ Create and schedule advertising campaigns for customers competing within the same industry to ensure the ads are never shown on the same page. ▪ Charge advertising customers based on the number of ad requests, clicks they want their ads to receive, and the page in which the ad displays.
Business Process Pipelines	▪ Customize order, targeting, and merchandising processes. ▪ Define and link the stages of a business process.

Security

Commerce Server supports the Secure Socket Layer (SSL) protocol to handle secure transactions. SSL provides the web server with the type of encryption and security required to complete commercial transactions.

Hardware And Software Installation Requirements

The ProLiant Sizer for E-Commerce Solutions tool can be used to determine optimal ProLiant server configurations for Commerce Server. These configurations are based on the minimum hardware and software requirements.

Commerce Server

Commerce Server requires the following minimum hardware configuration:

- 400MHz or faster Pentium-compatible processor
- 256MB of RAM
- 100MB of hard disk space
- CD-ROM drive
- Network adapter card
- VGA or Super VGA monitor
- Mouse or compatible pointing device

The volume of traffic on your website may dictate more stringent hardware requirements.

Before you install Commerce Server, you must install the following software on your server in the order listed:

- Microsoft Windows 2000 Server or Advanced Server
- The latest Windows 2000 Server service pack
- Microsoft Windows 2000 hotfixes
- An installation of SQL Server 2000 or SQL Server 7.0 must be available to Commerce Server on the same network.

Note

Commerce Server Direct Mailer must be installed on the same computer as SQL Server.

Administration Tools

The hardware requirements for Commerce Server Administration tools only are:

- 20MB of hard disk space
- VGA or Super VGA monitor set to at least 800 x 600 pixels resolution

Business Desk Client

The Business Desk client also requires a VGA or higher-resolution monitor set to at least 800 x 600 pixels resolution. Only 5MB of hard disk space is needed to install the application.

Commerce Server Business Desk client must be installed on one of the following platforms:

- Microsoft Windows 98
- Microsoft Windows Millennium Edition (ME)
- Microsoft Windows NT 4.0
- Windows 2000 Professional
- Windows 2000 Server
- Windows 2000 Advanced Server

In addition, Internet Explorer 5.5 and Microsoft Office Web Components or SQL Server Client Tools are also required. You must have a connection to the Internet to install the Business Desk client.

Commerce Server Business Desk client is required for all computers from which you want to access Business Desk, including the computer running Commerce Server. A high-speed connection to your Business Desk application is strongly recommended.

Note
A user with Administrator access rights must install the Business Desk client.

Commerce Server Solution Sites

A good starting point for building an e-commerce site is the Commerce Server Solution Sites. The Retail Solution Site and Supplier Solution Site incorporate the basic features needed to create a commercial website:

- **Merchandising** — Create personalized direct mail campaigns targeted and advertisements and discounts to increase sales and enhance the user experience.
- **Catalog display** — Display catalogs that visitors can easily search by category, full or partial names of products, or product attribute.
- **Customer service** — Customers can change their logon and password, view their order status and history, and change their profile information.
- **Order capture and receipt** — After ordering products, customers receive a receipt with a final total and order tracking number.

Retail Site

The Retail Solution Site includes business-to-consumer functionality for personalization, merchandising, catalog search, customer service, and business analytics. Customers have the ability to manage their own profile information and view their order status.

Supplier Site

The Supplier Solution Site is intended as a framework for designing and creating an e-commerce site that allows business-to-business (B2B) transactions such as purchase order and requisition handling. The site follows a B2B model in which products are offered to an established set of businesses. The Partner Services features enables you to designate an administrator to manage organizational information, purchase orders, and order status for that supplier company.

INTERNET The Solution Sites are not included with Commerce Server. Both sites can be downloaded from:

http://www.microsoft.com/commerceserver/solutionsites

Blank Site

Although the Retail and Supplier Solution Sites reduce the time and complexity required to build an e-commerce solution, the sites are general purpose, and their design may not satisfy the requirements of your company website. If strict and demanding performance requirements or unique architectural requirements are at the forefront, your company website might be better developed using the Blank Site.

The Blank Site, which is included with Commerce Server, allows you to customize your website to meet the specific needs of your business. You can begin with the Blank Site, which is a website that is essentially empty. The files contain very little content. However, the Blank Site also includes Commerce Server Business Desk and an empty set of databases on which the website will ultimately depend.

If your company has an existing website, you can convert it to take advantage of the features and functionality of Commerce Server. This requires unpacking the Blank Site, which converts and merges your existing files into the Blank Site. You will also have to customize the Global.asa file to allow your website to take advantage of the features of Commerce Server.

Global.asa File

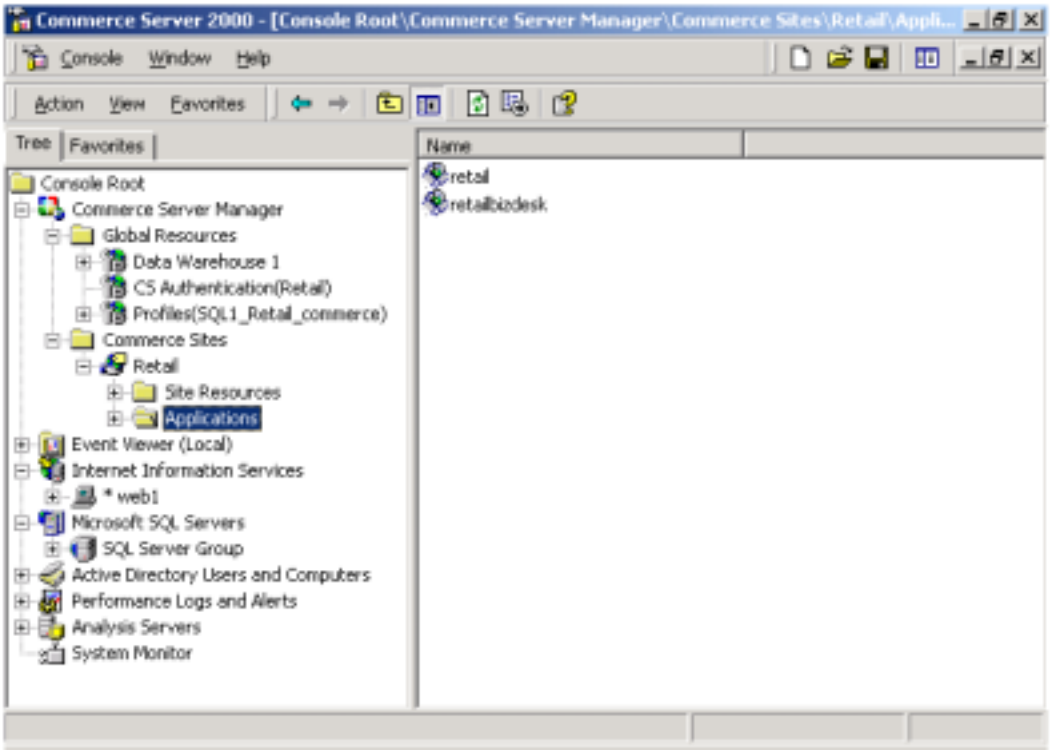
Typically when an application starts, it runs the Global.asa file. The Global.asa file contains code to create and initialize a large number of Commerce Server objects. The Internet Information Services (IIS) Application object stores references to those objects and services. As a result, any ASP in the application can access those objects and services through the Application object. Such access is often faster than creating and initializing, and then destroying, the requisite objects within each ASP page.

The types of processing that typically occur in the Global.asa file include:

- The creation and initialization of special purpose Commerce Server objects that can then be accessed from every ASP page on the site.
- The establishment of caching objects.
- The establishment of global information that is specific to the website.

One aspect of optimizing performance is to avoid unnecessary object creation. If you decide to use the Blank Site you will have to determine which objects are created and initialized in the Global.asa file and ultimately shared by all ASP pages.

Commerce Server Administration Tools



Commerce Server Manager and Commerce Server Site Packager are used to manage resources, sites, applications, and web servers.

Commerce Server Manager is accessed and manipulated through Component Object Model (COM) objects, so you can write a Windows Scripting Host (WSH) script or an ASP page to extend the administration environment. The Microsoft Management Console (MMC) hosts Commerce Server Manager.

Commerce Server Site Packager is a tool used to package a Commerce Server website and its applications and resources into a single file. You can package your site (including the IIS metabase settings), file system, resources from the Administration database, and SQL Server databases into a single file. You can also unpack the Commerce Server site (or portions of it) onto other computers.

Example

After you develop your website, you need to package and move your site from the development environment to the test environment. After the site has been tested, you want to deploy the website on a different computer. Site Packager can be used for each phase of the deployment until the website is moved to the production environment.

Preparing and Running Reports

Commerce Server provides several report definitions that you can customize and add to the Reports module. The following reports are accessed through the Commerce Server Business Desk.

- Advertising Reports provide information on campaign activity and user behavior in relation to website campaigns.
- Product Sales Reports allow you to review purchasing patterns, sales volume, revenue, and shopping basket activity.
- Web Usage Reports determine the entry paths and browsers visitors used to get to your website and the page request activity.
- User Reports track the number of unique visitors that visit your website and user registration information.
- Visit Reports determine which pages site users are most likely to visit.
- Diagnostic Reports allow you to review bandwidth consumption and diagnose problems users may be encountering when visiting your website.
- Query String Reports are used to analyze the content that users are seeking on your site.

  Example

  A financial firm wants to track the stocks in which users are most interested. This would require configuring a query string for the stock object and then running the Query String Report to analyze the query strings that you have configured for the system.
- Report Properties provide details about a report such as the name, description, type, date created, and the query that the report is based on. This is especially important if you are running a static report for a large data set, for example, running the Registered User Properties report for 10 million users. Static reports running on large data sets can require significant time to process.

HP Testing of Commerce Server

HP performed testing on two hardware configurations to determine the performance and scalability characteristics of Commerce Server and the Microsoft System Verification Test (SVT) sample site. The SVT is a Commerce Server business-to-consumer bookstore sample site.

One test was performed on a multi-server load-balanced (horizontal) configuration. The other test was conducted on a single-server multiprocessor (vertical) configuration.

Horizontal Testing

The horizontal configuration is based on DISA. In a horizontal configuration, Commerce Server is installed using Microsoft Network Load Balancing to distribute the user load across multiple web servers. The tests recorded the performance of SVT store operations scaled from one server up to eight load-balanced servers.

The following hardware platforms were used for horizontal testing:

- Eight ProLiant DL360 two-processor servers used as load generating clients.
- Eight ProLiant DL360 two-processor servers used for the Commerce Server web servers.
- One ProLiant DL580 two-processor server as the database server.

Number of Servers Used	Maximum Number of Concurrent Customers Supported	Relative Scaling (%)
1	2600	
2	5500	111.54%
2	5500	
4	11200	103.64%
4	11200	
8	21000	87.50%
1 to 4		330.77%
1 to 4		707.69%

Vertical Testing

The single-server configuration is based on a vertical scaling architecture. Commerce Server 2000 was installed on a multiprocessor web server to assess the load of each SVT store operation scaling the configuration for one to eight processors.

The following hardware platforms were used for vertical testing:

- Eight ProLiant DL360 two-processor servers used as load generating clients.
- One ProLiant 8500 eight-processor server as the Commerce Server web server.
- One ProLiant DL580 two-processor server as the database server.

Number of Servers Used	Maximum Number of Concurrent Customers Supported	Relative Scaling (%)
1P	2850	
2P	4200	47.37%
2P	4200	
4P	5800	38.10%
4P	5800	
8P	10100	74.14%
1 to 4P		103.51%
1 to 8P		254.39%

Testing Conclusions

The results from testing the two configurations indicated that Commerce Server scales almost linearly in a DISA horizontal architecture, but considerably less in a vertical symmetric multiprocessing architecture.

In addition, various software and hardware configurations were tested to determine which configurations impacted performance. Following is a summary of the results:

- The baseline testing conducted with Commerce Server used 1GB of RAM in the web servers in the DISA test configuration. Reducing the amount of memory to 0.5GB showed no performance degradation.
- IIS has three process isolation (memory space) options that can be configured for each application:
  - Low — all applications run under the Inetinfo.exe process
  - Medium — all application processes are pooled together
  - High — all application processes run under their own process memory space

  An abbreviated series of performance tests were run on both test configurations to observe the relative performance under each IIS process isolation option, which provided these results:
  - In the horizontal test configuration, optimum performance was observed while running under the medium option.
  - In the vertical test configuration, the high option showed the best performance. The expectation was to see the best performance while running in the low process isolation option because the application runs "in-process".

  Note
  During testing there can be anomalies, and attempts are made to understand them. This anomaly was noted and remains an open issue. HP will run more extensive IIS process isolation tests to understand and explain this anomaly.

- A significant decrease in performance was observed using the 1MB cache processors compared to the 2MB cache processors.
- A significant improvement in performance in the two-processor DISA configuration was noted compared to a one-processor DISA configuration.

INTERNET Testing details can be found in the "Commerce Server 2000 Performance Guide" at **http://www.hp.com/solutions/activeanswers**.

Learning Check

1. List the components within the Commerce Server architecture.

   .....

   .....

   .....

2. What is the purpose of a pipeline?

   .....

   .....

3. You want to collect information about users that visit your website. Which Commerce Server feature provides that functionality?

   .....

4. After you develop your website, you need to package and move your site from the development environment to the test environment. Which Commerce tool provides this functionality?

   .....

5. Which Commerce Server report would allow you to analyze the content users are seeking on your site?

   .....

6. Which process isolation option provided the best performance for the DISA configuration in HP testing?

   .....

Deploying Microsoft Content Management Server

Module 4

Objectives

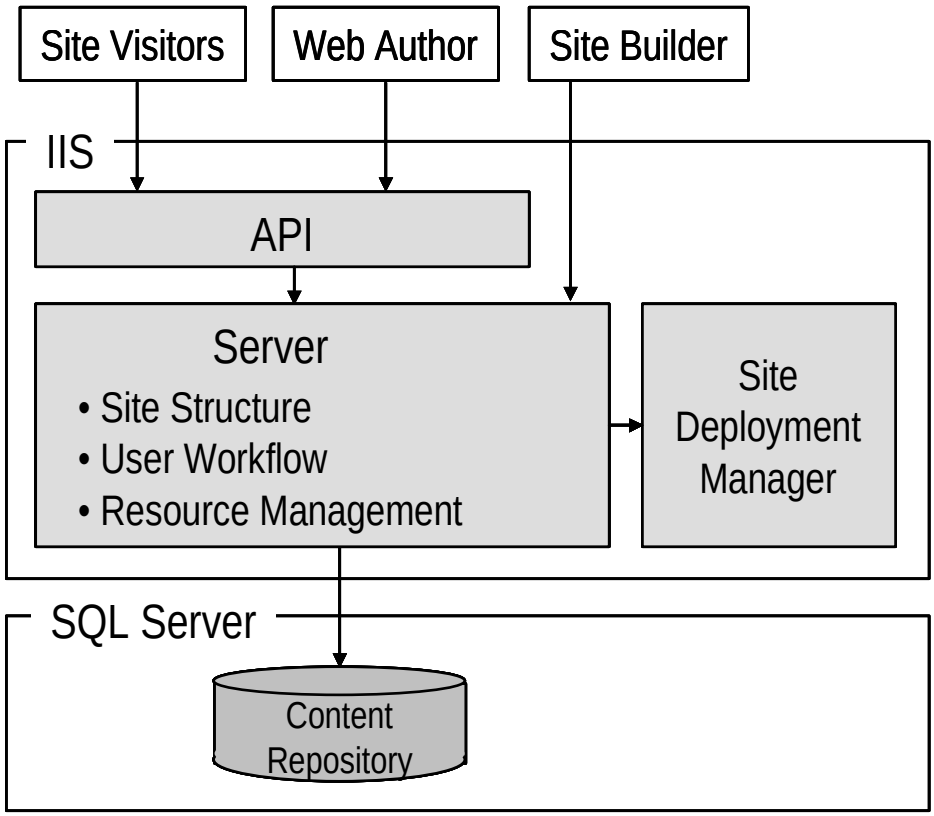
After completing this module, you should be able to:

- Describe the Microsoft Content Management Server (MSCMS) product and network architectures.
- Describe the functionality provided by MSCMS.
- Explain how each MSCMS role contributes to the workflow process to publish content to a website.
- Explain how the DISA framework can be used in MSCMS deployment scenarios.

MSCMS Architecture

Content Management Server is a distributed network application built on a relational SQL database. The components interact with each other and are used in each phase of the website implementation process.

MSCMS Product Architecture



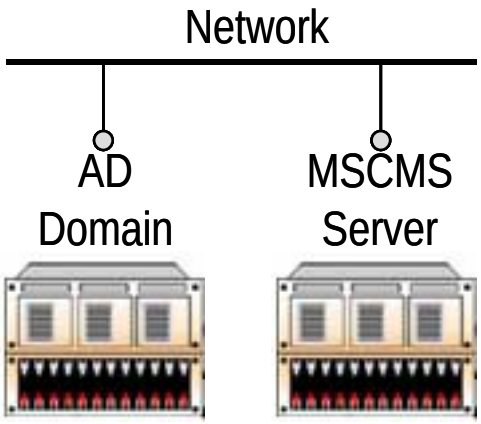
MSCMS is an application server that uses IIS as the web server. The preceding diagram shows the MSCMS product architecture.

The Content Repository stores all content, including templates, pages, and resources. It is accessible through the Publishing API, which allows ASP scripts to interact with the content stored in the SQL database.

When a site visitor requests a page, it is served dynamically from the Content Repository. Pages are collections of content and resources (for example, graphics, documents, and video) that are combined with template layouts at run time to produce site content.

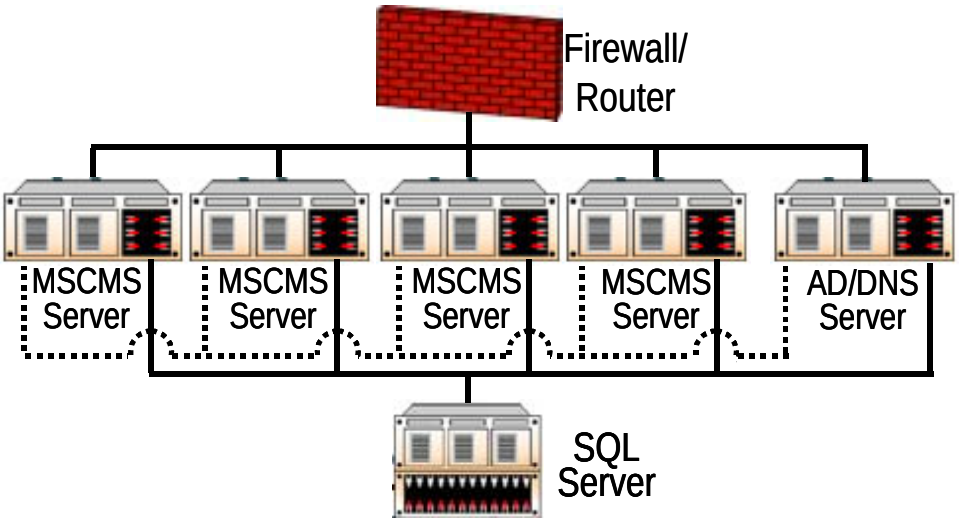
MSCMS Network Architecture

The MSCMS network architecture differs for internal and external deployments. MSCMS must have access to Active Directory Services (ADS) to authenticate users for website development and access. If more than one domain exists within the forest, a DNS server is required to access other ADS domains in the forest.



Typically in an intranet implementation, the people adding content and browsing the site are frequently the same users. This type of scenario does not necessitate multiple MSCMS servers or a router.

However, a public Internet site necessitates a configuration based on the HP DISA framework. A firewall is required to filter all traffic except traffic to Port 80. If clustered MSCMS servers are used, a load balancer directs traffic among the servers. The configuration also requires at least one domain controller for AD authentication and DNS services.



MSCMS Functionality

MSCMS is a comprehensive website environment that enables you to:

- Create, manage, and publish your own content.
- Quickly deploy scalable, dynamic sites.
- Deliver dynamic content for multiple audiences, devices, and purposes across enterprise websites.
- Deliver personalized content to customers, employees, and business partners.
- Support multiple languages within a single website.
- Analyze the structure and content of your site.

MSCMS Roles

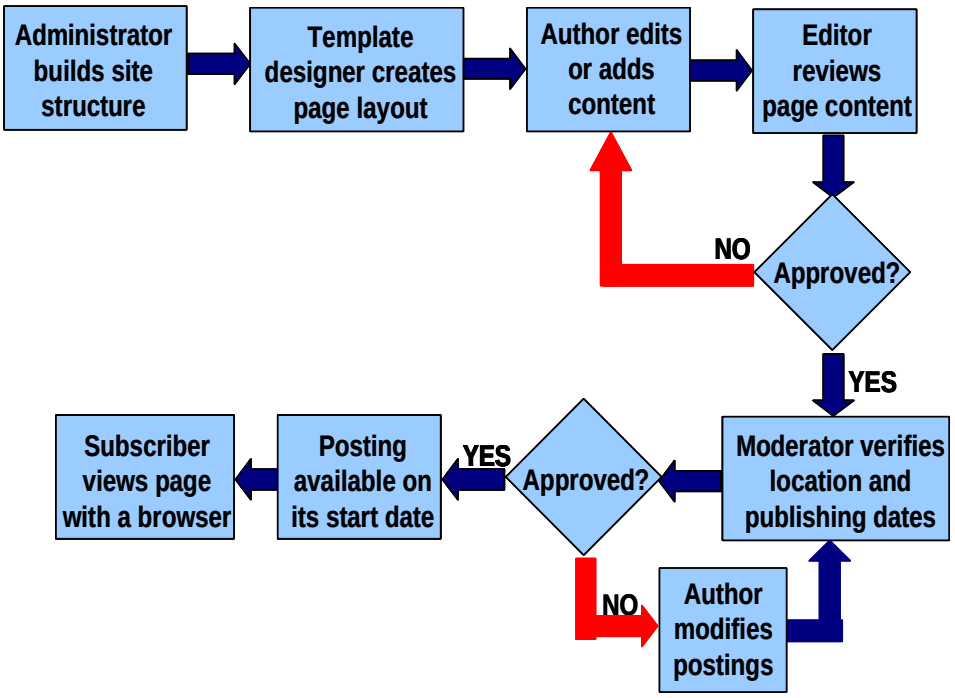
The workflow process is defined by roles. Each member in the workflow process must complete certain tasks to publish content published to a website. There are seven MSCMS roles:

- **Administrator** — Can do any task any other role can do. Administrators are solely responsible for setting up MSCMS hierarchies (folder, channel, gallery) and rights groups.
- **Template designer** — Creates the two types of MSCMS templates: page templates, which are used to create pages and navigation templates, which enable subscribers to navigate the channels of a site to view postings.
- **Author** — Creates and submits pages for approval by an editor. Some authors are associated with more than one folder. Authors can post their pages to a channel.
- **Editor** — Approves pages for publication. Editors are assigned to Site Builder folders and can post pages to a channel.
- **Moderator** — Approves postings to channels. This role ensures that the schedule and content are appropriate and relevant for each posting on a channel. Although a moderator can revise the schedule of a posting before approving it, they cannot revise the page content.
- **Resource manager** — Manages resource galleries. Resources are any file type that can be added to a page or template including Word documents, PDF files, ZIP files, image files, MP3s.
- **Subscriber** — Has access to channels to view pages published on a site. Users within a company are given access through their membership to subscriber groups. External users can log in to the site as guests and browse channels to which guest subscribers have been given access.

All users can be assigned to more than one role. Every role must at least be given subscriber rights to the containers in which they will work. This permits browsing to view content created on the website.

You can assign any Windows 2000 or LDAP account to the subscriber role, provided the account resides on a domain that is in a trust relationship with the machine hosting the server.

Workflow Process



MSCMS facilitates collaboration among a group of people in the website publishing process. This distributed authoring model works best when combined with a flexible workflow system.

The process begins with the creation of the site structure by the administrator. After the site structure is established, the template designer creates templates. All web pages in MSCMS are based on templates, which are stored in the template gallery.

The templates include areas called placeholders, which is where the author enters dynamic content. Placeholders can contain text, graphics, attachments, and links. After the placeholders are populated with content, the author saves the new page in one of the folders created by the administrator.

All pages in a site folder must be approved by an editor that has rights to the folder, or by an administrator, to continue through the publishing process. If no editor has been configured for a folder, authors can approve their own content.

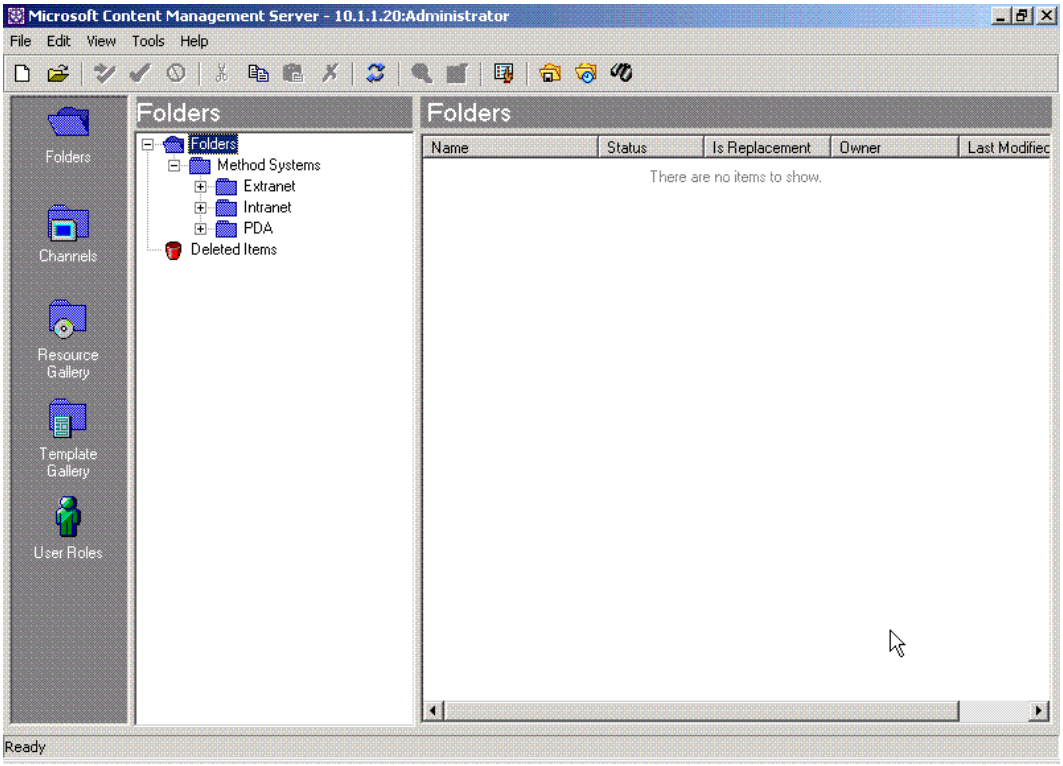
To publish content to the live site, the content must be posted in a channel in the site structure. The content is published according to a start and expiration date. The moderator, who approves or declines the publishing dates, reviews the content schedule. Postings are not available for viewing in a browser until approved by the moderator for the associated channel.

MSCMS Functional Components

Content Management Server has six main functional components:

- **Site Builder** — Used to create and manage the structure and resources of a website.
- **Web Author** — Used to publish and manage content.
- **Server Configuration Application** — Used to view and change the configuration values for the server and the database it references.
- **Site Deployment Manager** — Enables content, templates, and users to be moved between servers to share information across enterprise applications.
- **Site Stager** — Enables you to make an HTML or ASP copy of your website available for multiple platforms.
- **Content Connector** — Associates MSCMS roles with responsibilities and tools.

Site Builder



MSCMS Site Builder is used to build the application framework, set user roles and permissions, and build templates. It communicates with the Content Management Server over HTTP with or without SSL. Site Builder consists of several components:

- **Folders** — Contain pages in which content is stored.
- **Channels** — Show the navigational structure of a website and are used to store postings and control access to content. Postings contain pages to be published and include information such as the channel in which pages are published, the date the posting is available to subscribers, and the date the posting expires.
- **Resource Gallery** — Used to organize and manage access to resources, which include graphics, video documents, and text documents. For example, storing a company logo in the resource gallery means that every time you update it, all the pages or templates using that logo are automatically updated as well.
- **Template Gallery** — Manages access to page and navigation templates, which determine the design and structure of web pages. Navigation templates are used to generate links to pages, postings, and resources on your site.
- **User Roles** — Used to assign MSCMS permissions.

Web Author

Web Author provides authoring and publishing capabilities to users in a browser-based interface. Users enter content into placeholders by drag-and-drop from desktop applications such as Microsoft Word.

In MSCMS authors can create multiple content-related web pages using different template views. This allows users to share content across associated pages in instances where template families are used to target content to different audiences.

The user interface of the connected pages allows users to identify the associated web page views and to display their contents. Any changes made to a web page view are replicated across all shared placeholders in the associated web page views.

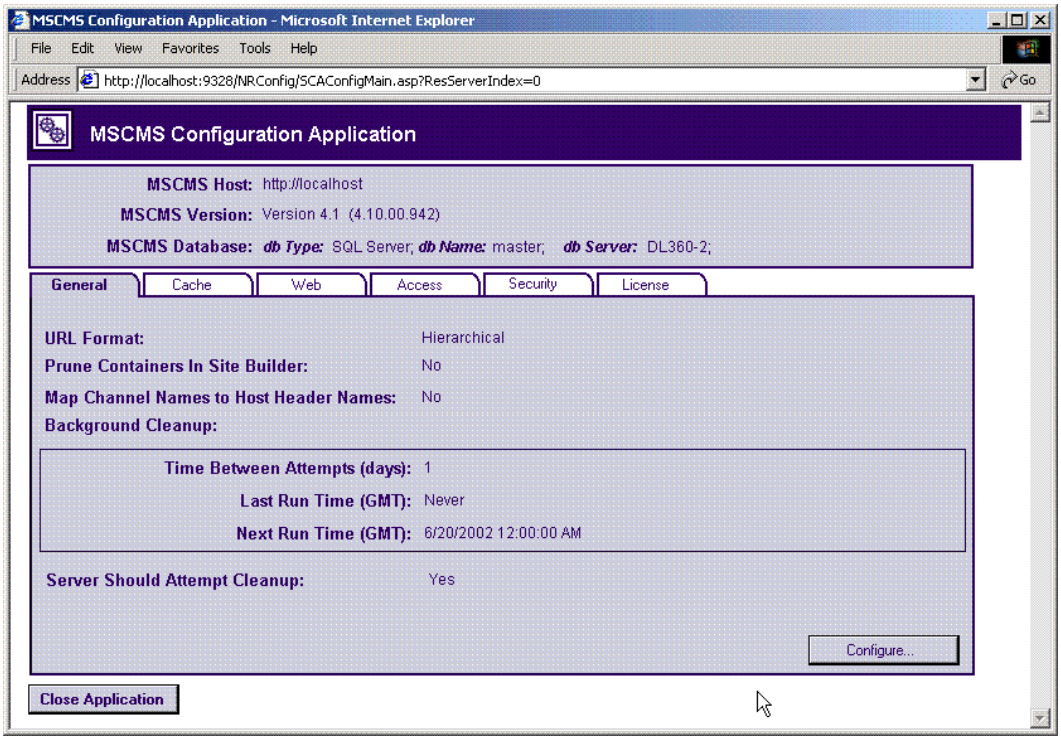
With Web Author, you can cross-post to multiple channels in which you have access rights so that the underlying content is shared.

Example

Outdated press releases are normally removed from the current site and moved to an archive area. Using the multiple posting feature of MSCMS, you can create new press releases based on the original one, but place them in different channels.

One purpose of multiple views is to share content among different sites on the same MSCMS server. Content can be in shared placeholders that belong to different templates, but share the same name. When the content of one placeholder is updated, all postings based on that placeholder will be updated at the same time.

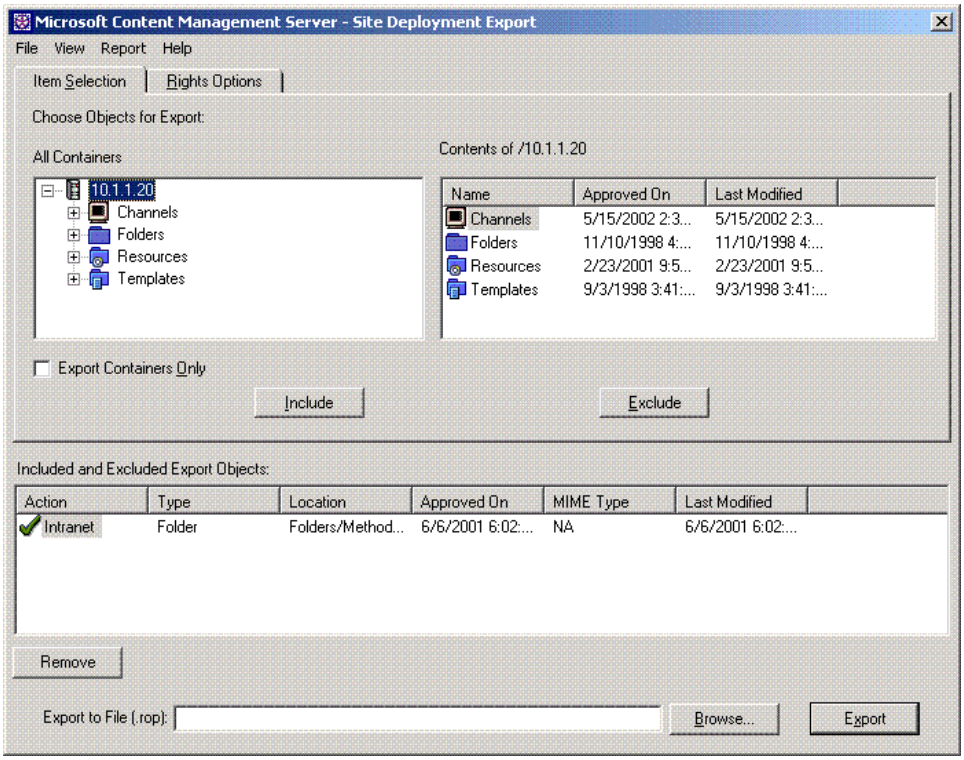
Server Configuration Application



The Server Configuration Application (SCA) allows you to perform administrative tasks on the Content Management Server website. These tasks are accessed through the following tabs that share the MSCMS database shown in the SCA:

- **General** — The default tab for the SCA. It allows you to:
  - Modify URL format (hierarchical or unique ID-based).
  - Manage viewing rights of each container.
  - Map channel names to host header names.
  - Perform background cleanup.
- **Cache** — Provides access to the settings for the disk cache and memory caches and includes a button to force the server to flush its memory cache.
- **Web** — Displays the status of web entry points for the MSCMS and provides access to the user interface used to add or remove entry points. This includes details for all visible virtual servers with an indication of their status with respect to Content Management Server.
- **Access** — Used to modify the list of currently supported NT domains and a list of supported Active Directory containers or Site Server Organizational Units if necessary.
- **Security** — Used to modify MSCMS System Account information, web cookie settings, and guest access options.
- **License** — Displays the current user's product license key information.

Site Deployment Manager



MSCMS Site Deployment Manager creates snapshots of site hierarchies and individual content objects and assembles them into an XML package for export. The XML packages contain the selected pages, as well as any dependant resources such as associated templates. Parts of a site can be exported from one Content Management Server to multiple target Content Management Servers, which allows you to separate your development and production environments.

Using the Site Deployment Manager to update content ensures that your web applications are available at all times. The Site Deployment Manager moves content in a transactional manner, while the site remains available to browsers.

Example

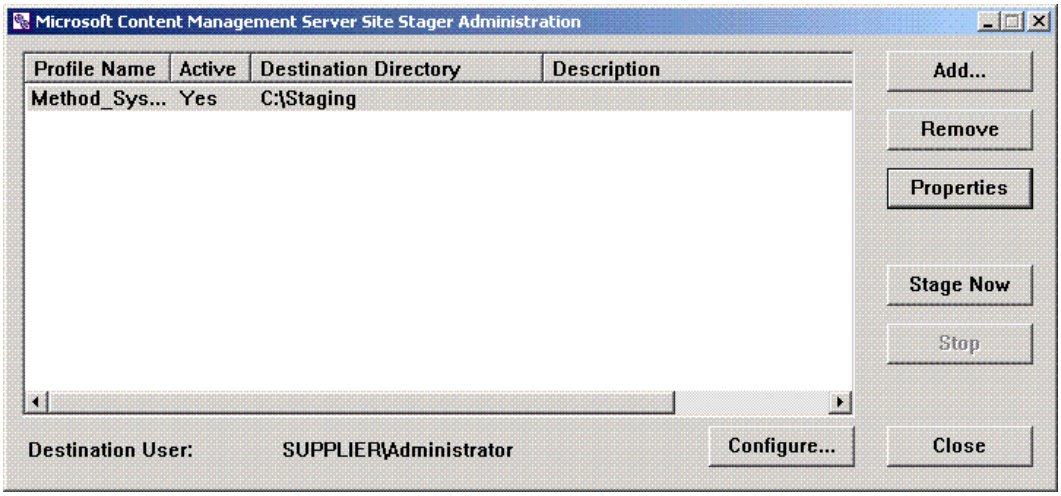
To transfer content from a development server to a read-only production server, isolate the channel structure of the site as well as authoring, editing, and moderating to the development server. Then use Site Deployment Manager to incrementally move content to the production server to synchronize the development and production environments.

Note

The Site Deployment Manager is designed for making incremental updates to a destination server. To copy an entire site, use a database replication utility.

When the Site Deployment Manager imports packages, it synchronizes the site hierarchy and contents on the target server to exactly match those of the source server. Objects on the destination system are locked while the import occurs to ensure data integrity.

Site Stager



MSCMS Site Stager creates an HTML or ASP snapshot of website content, navigation, and links for deployment on any server. The Site Stager is often used when content needs to be shared with a partner who may not be using Microsoft Content Management Server. If load balancing or server clustering is not required for your website implementation, the Site Stager can also be used to export a static version of the site to a failover server.

The HTML snapshots can contain script and can use any valid web file extension.

Staging can be scheduled to occur automatically. Because you can select the parts of an MSCMS site you want processed, you can set up more than one version of the web site to meet the specific information needs of various audiences, and automatically update these sites at regular intervals.

Staging is incremental. Only files that have changed since the last time the site was staged are exported. This allows you to maintain separate publishing and browsing environments, which frees resources for publishing and content management activity. Separating publishing and browsing also has security advantages because of reduced traffic on the web server.

Site Stager profiles determine the time at which staging is to occur, the frequency of staging, and which channels to stage. Site administrators create staging profiles that include the content they want to export as HTML. These profiles can be used to stage only part of a site.

On activation of a Site Stager profile, Site Stager navigates an MSCMS website, converting dynamic MSCMS pages into static HTML files. The HTML files, along with linked elements such as graphics, are organized in a directory structure that mirrors the channel hierarchy of the MSCMS website.

Site Stager requires the Task Scheduler service to be started. Make sure the Task Scheduler service is installed and configured to start automatically, using the Services control panel applet.

Content Connector

MSCMS integrates with Commerce Server through a feature called Content Connector. Content Connector allows site managers to personalize a site by targeting content to be shown to specific users on specific pages. Content Management Server also provides the ability to edit, enrich, and preview product pages, and to publish news releases or other content pages. These different pages can be targeted to specific users.

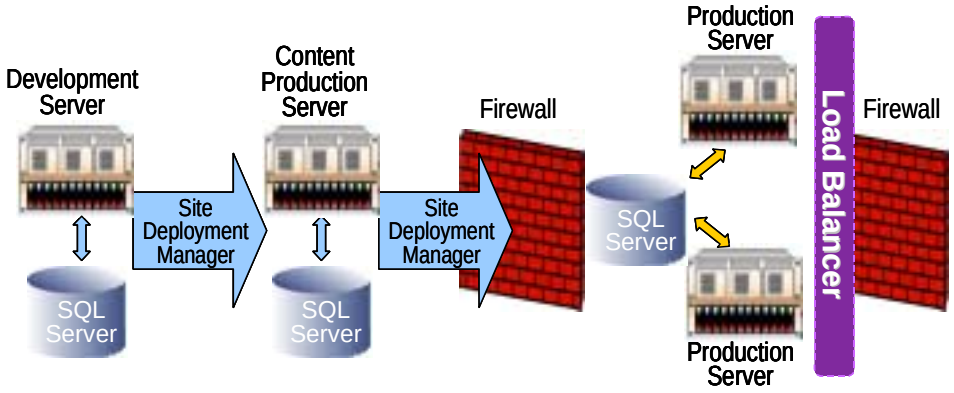
Content Connector includes add-ons such as additional Business Desk modules, extra MSCMS page templates, extensions to the Web Author, and include files. These Content Connector add-ons make use of Commerce Server objects.

Content Connector also keeps a log of user activities to assist in the tuning of the site for better personalization. Reports are created from the log that tracks the content viewed by website visitors. This information enables you to adjust content and context profiling accordingly.

Deployment Scenarios

This section defines the physical configuration for an Internet deployment and an intranet deployment. An Internet deployment is divided into three separate environments: development, content production (test), and production. The environment is protected by a firewall between the servers and the Internet, and between the servers in this environment and the internal corporate network.

Internet Deployment



In the development environment, each site developer uses a separate instance of Content Management Server to access the same database. When functional code is checked in, it is run on the development server.

When the site is ready to be deployed in the content production environment, SQL database transfer (or backup and restore) is used to move the entire site. For any subsequent incremental changes, Site Deployment Manager is used to move site structure, templates, and content.

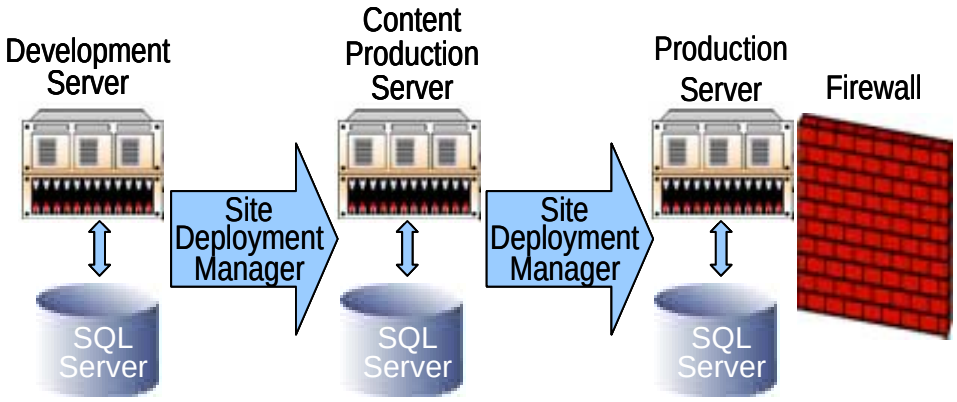
In the content production environment, content contributors using the Web Author add content to the site. It is also the final staging or test environment before the site is live. This entire environment is behind the corporate firewall or firewalls.

The production environment serves the site to the outside world. Production servers are clustered. One IP address is available to the Internet, and network load balancer directs traffic among the servers to ensure performance and availability.

Example

If a site has frequent feature additions or template changes, site building and testing may occur on a development server; content development and testing occur on a separate server; and the site is served and browsed on clustered production servers with a load balancer to provide scalability, performance, and availability. Website modifications are completed on the development server without affecting content production. To keep the servers synchronized, production servers are used as read-only servers.

Intranet Deployment



In a typical intranet deployment, there are two environments, development and production. Site building and testing are done on the development server before any changes are moved to the production server. The production environment is used for creating content and browsing the intranet site. Both environments are within the corporate network, protected by a firewall.

Transferring content from a development server to a production server can also serve as a failover or replication solution. This deployment scenario is used in small deployments where the traffic does not warrant load-balancing hardware or software.

Example

When the website is ready to be deployed into the content production environment, use SQL database transfer (or backup and restore) to move the entire site. For any subsequent incremental changes, use the Site Deployment Manager to move site structure, templates, and content.

Learning Check

1. List the six main functional components of Microsoft Content Management Server.

   .....

   .....

   .....

   .....

   .....

   .....

2. A human resources staff member, who is assigned the role of moderator, needs to make a change to a job posting but is unable to access the page content. What is the problem?

   .....

   .....

   .....

   .....

3. You need to add several new white papers to your website. Which Content Management Server component would you use?

   .....

4. When a website is ready to be deployed into the content production environment, what application would you use to move the entire site?

   .....

   .....

Managing Website Implementations

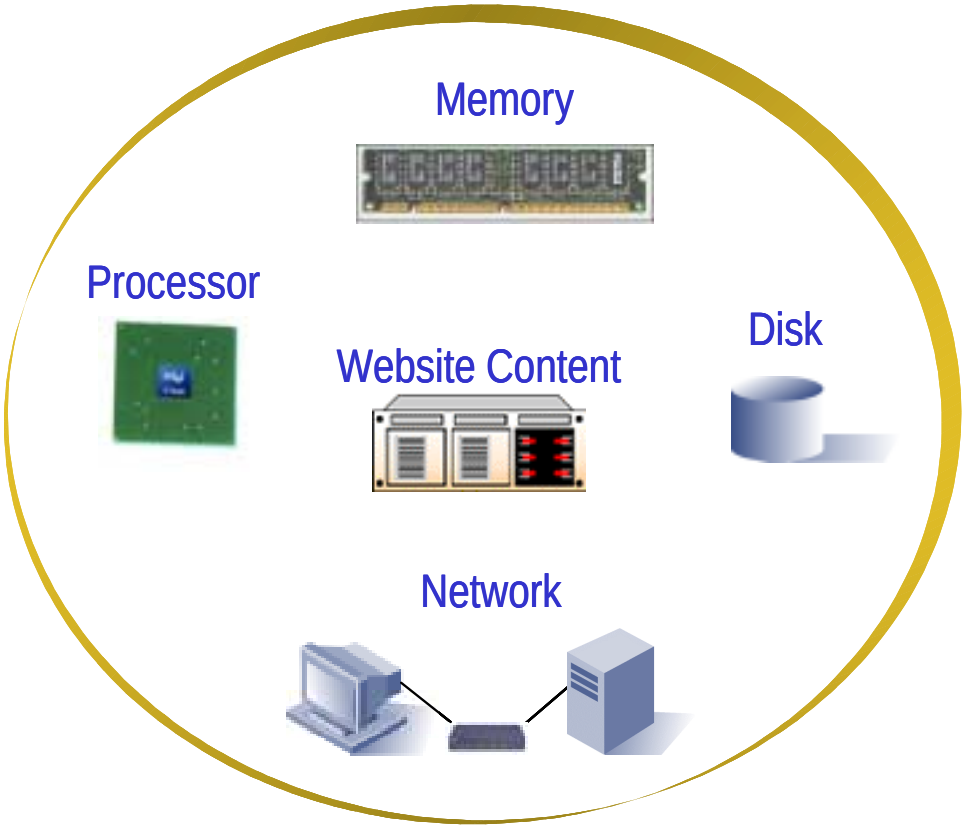
Module 5

Objectives

After completing this module, you should be able to:

- Describe the factors that can affect web server performance and list potential web server bottlenecks.
- List the goals of capacity planning, the main capacity planning metrics, and key performance measurements.
- List and explain the HP recommendations for capacity planning in relation to sizing the processor, memory, disks, and network.
- Monitor server subsystems to identify potential bottlenecks.
- Identify the counters used to collect data to analyze web server performance.
- Tune the processor, memory, disk, and network subsystems for optimal web server performance.

Factors That Affect Web Server Performance



Web administrators need to determine the hardware configuration of the web server based on assumptions about its use and future expansion needs.

System throughput problems, or bottlenecks, usually occur when the demand for resources exceeds supply. Isolating performance problems starts with determining how users, web server applications, and the operating system interact with each resource.

The four main hardware components that affect web server performance are:

- Processor
- Memory
- Network
- Disk

The type of content served by the website also affects performance.

Processor Bottlenecks

You can identify a processor bottleneck by:

- A long, sustained processor queue.
- High use rates on one or more processors.
- Connections being blocked or rejected.

Congested Queue

A processor bottleneck occurs when one or more processes consume nearly all the time of all processors on the computer. In a bottleneck, process threads that are ready to be executed must wait in a queue for processor time. All other activity comes to a halt until the queue is cleared.

If the work in the queue is not distributed to the other processors on a multiprocessor system, a bottleneck can occur when a single processor is fully loaded. A sustained processor queue length of two or more threads typically indicates a processor bottleneck.

Visitors may experience a delay in response time when the website has dynamic content. With dynamic content, the processor is usually the bottleneck. For example, when a server receives a request for a dynamic page, it can process server-side scripts to build the web page that is sent to the client browser as well as call COM components that perform a variety of tasks, such as connecting to a database or processing business logic.

Constant Activity

Sustained high rates of processor activity, which do not leave capacity to handle peak loads, are also associated with processor bottlenecks. Processor activity itself only indicates that the resource is being used, not that maximum use of the resource is a problem. However, a long, sustained queue indicates that threads are being kept waiting because a processor cannot handle the load assigned to it.

Blocked Connections

Each connection that an IIS request establishes consumes some processor time. The network adapter card interrupts the processor to signal that a client has requested a connection. Further processing is required to establish and maintain the connection. When the connection is closed, processing is required to delete the structures that serviced the connection. Each time a connection is established, the load on the server increases.

Memory Bottlenecks

System performance is affected by the way in which various applications use memory. Frequently used information should not be sent to disk because this increases the amount of disk activity, which slows the server and degrades performance.

To better understand the terms used to assess memory bottlenecks, consider the following example.

Example

The largest element of activity in Windows 2000 Server is a process. The physical RAM available to a process is called its working set. If the process cannot store all of its code and frequently used data in physical memory, some of the information must be stored elsewhere, usually on disk (as virtual memory). This type of memory region is designated as *nonpageable*. causes an in,.

A memory bottleneck can sometimes look like a processor or disk bottleneck. If the system does not have enough physical memory to store the code and data that are needed, the processor spends substantial time paging. Before adding or upgrading processors or disks, examine memory usage in the server.

Web Server Bottlenecks

Within each process, Windows 2000 uses threads to accomplish particular tasks. The memory occupied by threads is part of the working set of the process, except for some special threads that run in nonpageable memory. Threads that service website connections are not pageable. If too many nonpaged threads are created, the system runs out of nonpaged pool memory and a bottleneck occurs.

When the amount of available memory falls below 4 MB, the system attempts to provide more available bytes by taking memory away from the working sets of processes. This strategy increases the rate of page faults because each process must now retrieve data either from disk or from elsewhere in memory.

When the rate of page faults for a particular process rises, the system attempts to expand the working set of that process to lower the page fault rate. As the system tries to maintain this balance, the size of the working set of each process fluctuates accordingly.

The IIS Object Cache stores objects that are costly to retrieve and frequently reused by the service, such as file handles and file directory listings. It is a cache maintained by the IIS 5.0 service, part of the working set of the IIS 5.0 process, Inetinfo.exe, and can be paged to disk.

Cache flushes can affect the performance of the IIS Object Cache. An internal timer regulates cache flushes. The timer activates the object-cache scavenger, which deletes expired objects. Objects are flushed from the cache if they change, or if they time out before they are reused.

A hit is recorded when requested files are found in the cache. A miss or fault is recorded when requested files are not found. Misses and faults indicate how often

the system needs to do extra work to retrieve files from somewhere other than the cache.

If a high rate of cache flushes is associated with elevated cache misses and page faults, it is possible that the cache is being flushed too frequently. A high rate of cache faults can indicate a memory shortage.

Database Server Bottlenecks

If the user connections option is left at the default value of zero, SQL Server will dynamically adjust the number of concurrent user connections to the maximum value. It is preferable to leave this option at the default value.

You can set the user connections option to another value to prevent SQL Server from overloading from too many user connections. Estimate the maximum number of concurrent user connections and add five to that value. User connections that exceed the maximum allowed are rejected.

If you set the user connections option to a value other than the default, SQL Server will reserve approximately 40KB of memory for each reserved user connection, regardless of whether the connection has been established. This can cause a memory bottleneck on the database server.

Network Bottlenecks

The primary functions of a website are to establish connections for its clients, receive and interpret requests, and deliver files. The pace at which these vital functions are performed depends on two factors: the effective bandwidth of the link between the server and the network, and the capacity of this link and the server to support network resources.

Bandwidth is measured in several different ways:

- The rate at which bytes are transferred to and from the server.
- The rate at which data packages are sent by the server. Data packages include frames, packets, segments, and datagrams.
- The rate at which files are sent and received by the server.

Bandwidth shortages can be detected by monitoring the success and failure of connections established and rejected by TCP. With ample bandwidth, the server can establish and serve connections before they time out. If it is not sufficient, the connections fail.

Disk Bottlenecks

The disk subsystem can easily become the bottleneck on any system. On the front-end web server, the disk utilization should be low because the content and images for a page should fit well within the file cache. The primary disk activity is the log file. Under ordinary circumstances, disk activity, other than that generated by logging, serves as an indicator of issues in other areas.

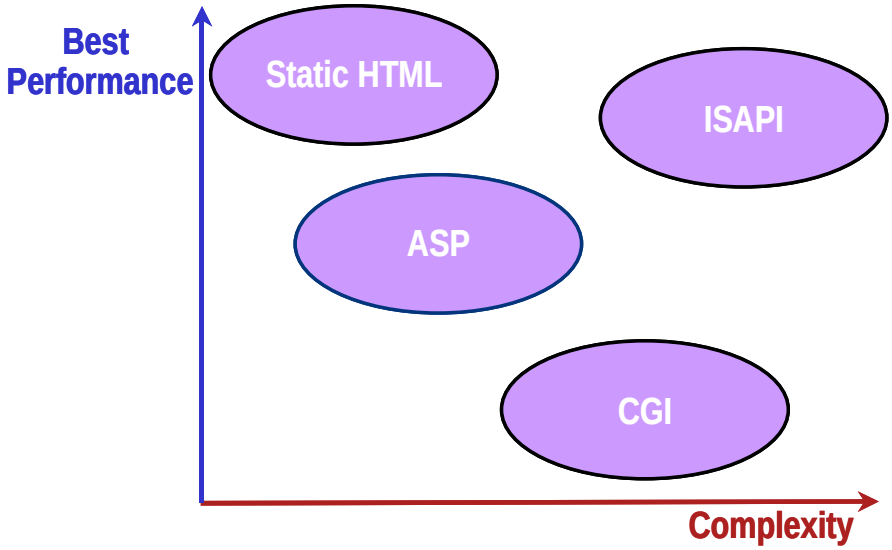
Example

If your server needs more RAM, you will see an increase in disk activity due to a rise in hard page faults. However, there will also be much disk activity if your server houses a database or your users request many different pages.

If you do not have anything that is obviously disk-intensive on your server, but you see an increase disk activity, check RAM use to ensure you have enough memory.

The database server makes extensive use of the physical disk subsystem. Planning and calibrating this subsystem is the key component of a fast data resource server.

How Website Content Affects Performance



As the content becomes more complex, web servers have to do more work per visitor, which lowers site capacity. Minimizing database use and dynamic content, as well as using simpler HTML pages, can increase site capacity. However, web servers should be optimized for running both static and dynamic content.

Dynamic content includes:

- Common Gateway Interface (CGI)
- Active Server Pages (ASP)
- Internet Server Application Program Interface (ISAPI)

CGI

CGI scripting provides dynamic generation of HTML on the web server side. With CGI scripting, web servers can run programs that generate the HTML files to be sent to the web client rather than precomposed, static information.

ASP

Although CGI provides a degree of dynamism on the web server side, ASP combines HTML, server-side scripts, and reusable components to create dynamic and powerful web-based business solutions. When a client requests an ASP file, the server parses the file and processes the scripts that are designated to be run on the server. These scripts can access databases, start out-of-process applications, make COM calls, write JavaScript or HTML that will later be interpreted on the Web client side, or write to the server disk.

ISAPI

Because ASP is an interpreted language that must be dynamically parsed by the web server each time the page is executed, the overhead limits the performance. The solution is to translate critical ASP pages into compiled ISAPI extensions and to convert relatively static ASP pages into HTML.

Capacity Planning

The goal of capacity planning is to determine the computing and network resources required to match the needs of the business.

The capacity planning process involves formulating a set of “what if” questions related to the business and computing environment, and subsequently providing the appropriate answers to these questions.

One of the most critical tasks in capacity planning is predicting whether system saturation will occur. This includes determining the maximum user load and throughput of the system being tested.

Proper capacity planning can:

- Reduce overall cost of ownership of a computer system.
- Identify potential bottlenecks before they occur.
- Improve user response time.
- Help plan for upgrades.

Capacity Planning Metrics

The three main capacity planning metrics are:

- **Requests per second** — The rate at which the web server services requests from all clients. A benchmark program such as the Web Application Stress Tool (WAST) can be used to determine this value.
- **Throughput (megabits/sec)** — The rate at which the server can process data. This value tells you how much data the server is moving to the clients. A benchmark program is used to determine this value.
- **Processor utilization** — How much work is being done by the web server. In a Windows 2000 environment, this value is measured by the % Processor Time counter in System Monitor.

Network and disk performance can also be measured; however, HP testing shows that these items do not significantly affect server performance.

A web server reaches its upper limit of performance when its processor (or processors) reaches 100% utilization.

Performance should be measured with different types of loads (static, dynamic, or a combination). A common load ratio used in web server testing is 80% static and 20% dynamic.

Web Server Hardware Capacity Planning Guidelines

To minimize bottlenecks, HP recommends the following capacity planning guidelines:

- **Processor** — Have enough processing power to handle the expected user requests. Remember that dynamic content requires more processing power and memory than static content.

  If a processor consistently exceeds 80%, upgrade to a faster processor, add a second processor to the server, or add more servers by setting up a multiserver environment.

- **Memory** — Have enough memory in the system to reduce or eliminate paging. If there is excessive paging on the disk subsystem and the page fault rate indicated by the Pages/Sec counter in Performance Monitor is high, add memory to the server.

  Striping a paging file across separate physical drives improves paging file performance. Typically, that the paging file on the system drive should be at least twice the size of physical memory so that if a crash occurs the system can write the entire contents of RAM to disk. This *core dump* can help you recover data and determine the cause of the crash.

- **Disk** — Plan for enough disk storage to support additional content for at least two years. When using an array controller, configure a RAID implementation such as HP Advanced Data Guarding (ADG) for data protection. ADG reduces the risk of an array failure and requires less storage capacity overhead than RAID-1.

INTERNET For more information on ADG, see the HP Reference Library document, "New Advanced RAID Level for Today's Larger Storage Capacities: Advanced Data Guarding" at **http://www.hp.com/support**

- **Network** — Ensure that network bandwidth is sufficient to handle user requests. Depending on your network infrastructure, adding a second NIC to the web server or segmenting the network will reduce network congestion.

Monitoring Subsystems

Monitoring the performance of the web server allows you to recognize problematic trends as they develop and take steps to prevent them from turning into emergencies. Monitoring also helps you decide when to upgrade your hardware, and whether upgrades are improving the performance of your server appropriately. Also, while some server problems are obvious, some develop over time. It is best to have a baseline against which to judge and compare.

Planning for and configuring your monitoring tools are the initial steps for monitoring your sites. Use System Monitor to decide how you should best tune your server.

Processor Monitoring

To prevent processor bottlenecks, ensure that a lengthy processor queue is not forming when you serve large numbers of connections. You can usually avoid a bottleneck during peak time by setting the connection limit to twice the average value of Current Connections.

If the processor regularly becomes a bottleneck when servicing large numbers of connections, consider upgrading or adding processors, or limiting the maximum number of connections on the server. Limiting connections can cause the server to block or reject connections, but it helps to ensure that accepted connections are processed promptly.

Memory Monitoring

Use the following recommendations to verify that the memory requirements of a web server have been met:

- Ensure that your server has enough RAM to keep the entire Inetinfo process in memory at all times. Web services run in a pageable user-mode process called Inetinfo.exe. When a process is pageable and enough free memory is available, the system can remove part or all of the Inetinfo process from RAM and write it to disk. If part of the Inetinfo process is paged to disk, the performance of the website suffers.
- Web server applications read sequentially. Sequential reading uses a Cache Manager feature called a read ahead. This occurs when the predictive algorithms of Cache Manager detects sequential reading and begin to read larger blocks of data in each read operation. Read aheads can provide a significant performance boost to a process.
- The File System Cache is particularly useful for servers of static web pages because web pages tend to be used in repeated, predictable patterns.
- In general, when memory is scarce the system trims the cache; and when memory is ample, the system enlarges the cache. If performance degrades when the cache is large, consider adding RAM. If performance degrades when the cache is small, consider leaving objects in the cache longer before you allow them to be flushed.

Disk Monitoring

One of the most effective monitoring tools for websites with static content is the web server log file. You should analyze this file periodically to determine which content is requested most frequently (at least 50-80%) and make adjustments to the web server disk configuration as needed.

Network Monitoring

You can increase existing bandwidth by increasing the length of the connection queues. Requests for connections to web services are held in queues until the service is available to respond to the request. A separate queue exists for each web service, but all queues have the same maximum size. By default, each queue can hold up to 15 connection requests. If the queue to a service is full, any new connection requests are rejected.

To ensure optimal bandwidth, you can also verify that HTTP Keep-Alives is enabled. HTTP Keep-Alives maintains a connection even after the initial request is complete. HTTP Keep-Alives is enabled by default, but can be disabled.

To disable HTTP Keep-Alives:

1. In the Internet Services Manager, right-click a site, then select *Properties*
2. Select the Performance tab
3. Deselect the *HTTP Keep-Alives Enabled* check box.

Useful System Monitor Counters

System Monitor uses performance logs and alerts to collect data called performance counters. A performance counter is a data item associated with a performance object. For each counter you select, System Monitor presents a value corresponding to a particular aspect of the performance defined for the performance object.

Performance counters allow you to monitor specific aspects of work performed on your site by a system or service. System Monitor groups counters by object type.

- **Active Server Pages → Request execution time** — The time spent processing the actual script. If this counter increases dramatically or the time goes over one second, the system is working beyond its capacity. Web pages should be designed to process in less than one second.
- **Active Server Pages → Request wait time** — The amount of time a new request waits in the queue before it begins processing. This value should be very close to zero (less than 100 milliseconds) because you do not want users to experience long wait times to process their checkout transactions.
- **Active Server Pages → Requests executing** — Number of requests running simultaneously. There should only be one request executing at a time.
- **Active Server Pages → Requests queued** — The actual throughput of the dynamic pages. Includes successful and failed requests.
- **Active Server Pages → Requests per second** — The actual throughput of the dynamic pages. Includes successful and failed requests.
- **CS2000: Marketing and Catalog → LRUCache: Cache size** — The size of all the caches in the system combined in bytes. This size should be increased in the Global.asa to maximize the cache hit rate.
- **CS2000: Marketing and Catalog → LRUCache: Flushes/sec** — The number of items flushed out of the cache per second to accommodate new items added to the cache. The Least Recently Used Cache (LRUCache) flushes items from its cache if it is programmatically requested or if the number of elements for the cache exceeds the maximum size set in the Global.asa by 10%.
- **CS2000: Marketing and Catalog → LRUCache: Hits/sec** — The rate of successful requests that are served from the LRUCache in memory.
- **CS2000: Marketing and Catalog → LRUCache: Misses/sec** — The rate of unsuccessful requests that are served from the LRUCache in memory. This can also be viewed as the rate that new items are added to the cache.

- **Memory → Available Bytes** — Measures the amount of memory available to the server. If the available megabytes goes too low then the system will start paging. The absolute minimum number is four, but it is a good idea to keep headroom on a server to account for peaks.
- **Memory → Page faults/sec** — Tracks the number of times the system cannot find a page in memory at the requested location.
- **Memory → Pages/sec** — Measures the actual memory requests that are made to the hard disk. This is a key indicator of a lack of memory resources or a poorly laid out solution.
- **Network Segment → Bytes received per second, Bytes sent per second** — Tracks the number of bytes received or sent per second on a network segment Monitor this value to ensure that it does not reach the capacity for the network card. If this value approaches the capacity of the network, then a higher bandwidth network might be necessary.
- **Physical Disk → % Disk Time, Disk Reads/sec, Disk Writes/sec** — Three counters that track the activity in the disk subsystem. The disk subsystem can easily become the bottleneck on any system. On the front-end web server, the disk utilization should be low because the content and images for a page should fit well within the file cache. The primary disk activity is the log file.
- **Processor → % Processor Time** — Shows the percentage of time that the processor is executing application or operating system processes other than the Idle thread, which consumes cycles when no other threads are ready to run.
- **SQL Server:Databases → Transactions/sec** — Indicates for activity in the back-end SQL Server.
- **SQL Server:Databases → Compiles/sec** or **Recompiles/sec** — Indicates the efficiency of the queries that SQL Server is executing. Reducing this number is key to reducing the processor load on the SQL Server.
- **System → Context Switches/sec** — Represents the number of times the system swaps from one thread to another. If this counter goes above 5000 per processor, it indicates poor SMP scalability for the server and application. The components of Windows 2000 and Commerce Server 2000 were designed to scale well.
- **Web Service → Get Requests/sec** — Represents the sum of the dynamic pages and static pages. This is the key counter for throughput.

Optimizing Web Server Performance

Websites based on a DISA configuration include multiple tiers, which can introduce many types of performance problems. Optimizing the performance of a website mainly consists of finding and removing bottlenecks at each tier in the configuration.

Processor Optimization

IIS runs in a set of multithreaded processes designed for efficient scaling on single-processor and multiprocessor systems. The relationship between threads, connections, and requests is complex because IIS uses the worker thread model. Instead of dedicating a thread to each connection or request, IIS dedicates one pool of threads, the worker threads, to the task of accepting and monitoring all connections. This frees other threads to do the remaining work of the application, such as:

- Authenticating users
- Parsing client requests
- Locating, opening, and transmitting files
- Managing internal data structures

Following are recommendations for optimizing the processor subsystem:

- **Monitor context switches** — Observe the patterns of context switches over time, which indicate that the kernel has switched the processor from one thread to another. A context switch occurs each time a new thread runs, and each time one thread takes over from another. A large number of threads is likely to increase the number of context switches. Although content switches allow multiple threads to share the processor, they also interrupt the processor and might interfere with its performance.
- **Upgrade the L2 cache** — When adding or upgrading processors, choose processors with a large secondary (L2) cache. File server applications, such as IIS, benefit from a large processor cache because their instruction paths involve many different components. A large processor cache (2 MB or more if external, up to the maximum available if on the processor chip) is recommended to improve performance on active servers running IIS.
- **Add network adapters** — If you are administering a multiprocessor system that does not distribute interrupts symmetrically, you can improve the distribution of the processor work load by adding network adapters so that there is one adapter for every processor.

  However, if one of the processors is nearly always active (% Processor Time = 100) and more than half of its time is spent servicing deferred procedure calls (DPCs) where % DPC Time > 50, then adding an adapter is likely to improve system performance, as long as the available network bandwidth is not already saturated.

DPCs are similar to interrupts except that they have a lower Interrupt Request Level. Unlike interrupts, DPCs can be delayed, allowing the processor to complete higher priority work.

Load can be distributed across multiple network adapters that service the same IP address by using HP Network Adapter Teaming. HP Network Adapter Teaming provides fault tolerance and load balancing across a team of two or more network adapters.

INTERNET For more information on HP Network Adapter Teaming refer to the whitepaper in the HP Reference Library "Compaq Network Adapter Teaming" at **http://www.hp.com/support**

- **Upgrade network cards** — HP 10/100 network adapters support interrupt moderation. Interrupt moderation prevents the processor from being overwhelmed by bursts of interrupts. Under normal operation, the adapter generates an interrupt every time a frame is received. Reducing the number of interrupts improves processor use.
- **Limit connections** — If you cannot upgrade or add processors, consider reducing the maximum number of connections that each web service accepts. Limiting connections can result in connections being blocked or rejected, but it helps ensure that accepted connections are processed promptly.
- **Redesign the website** — You can improve performance and reduce the processor work load by optimizing database use, optimizing the design of ASP pages, calling compiled components from scripts in ASP pages, using ISAPI instead of ASP, substituting static web pages for dynamic pages, eliminating the use of SSL except where it is necessary, moving trusted out-of-process applications into the Inetinfo process, and eliminating large bitmapped images or optimizing them to reduce their size. It is also advisable to set long expirations on infrequently modified static files.
- **Increase threads per processor** — By default, the IIS process creates up to four threads per processor. For most systems, this tuning is sufficient to maintain the optimum number of threads, but you can change the maximum number of threads per processor if your system requires it. If nearly all threads seem to be busy, but the processors are not always active, consider increasing the maximum number of threads per processor.

Memory Optimization

The primary purpose of monitoring memory in a web server is to ensure that the system has enough memory and is using it efficiently. By default, the system reserves about 50% of physical memory for the File System Cache, but the system trims the cache if it is running out of memory.

Web servers benefit from ample physical memory. Generally, the more memory you add, the more the servers use and the better they perform. At least 512MB of memory is recommended.

Following are some suggestions for optimizing memory performance:

- Keep related web files on the same logical partitions of a disk. Keeping files together improves the performance of the File System Cache. Also, defragment your disks. Even well organized files take more time to retrieve if they are fragmented.
- Have enough physical memory to hold all static web pages. However, if pages must be retrieved from disk, use mirroring or striping to make reading from disk sets faster. In some cases, a caching disk controller may help.

  Disk caching affects performance by reducing disk access and assisting in disk I/O request, sorting, and queuing. HP array controllers feature a configurable disk cache that supports read-only caching, write caching, or a combination of read-only and write caching.
- Add paging files and increase the size of the ones you have.
- Add memory to improve performance if the system cannot lower the page fault rate to an acceptable level.
- Lengthen the period that an unused object can remain in the cache with the ObjectCacheTTL setting in the registry.
- Limit the number of connections on the server to alleviate the memory shortage because some physical memory is consumed by the data structures the system uses to keep track of connections.
- Add memory to increase the size of the cache and improve its performance. If Cache Hits and Cache Hits % are low, or if Cache Misses is high, the cache could be too small to function effectively. Cache performance is judged by how often objects sought in the cache are found. Frequent cache misses result in disk I/O and decreased performance. A value of 80% to 90% for Cache Hits % is considered to be excellent for sites with many static files.
- Most processors can handle large numbers of software page faults without consequence. If the number of hard page faults is high, there may be too much memory dedicated to the caches, which does not leave enough memory for the rest of the system. Sustained hard page faults of over five per second are a key indicator of not having enough RAM. Increase the amount of RAM on the server or lower the cache sizes.

Network Optimization

The simplest measure of the effective bandwidth of a server is the rate at which the server sends and receives data. The components that collect data each reside in different Open Systems Interconnectivity (OSI) layers:

- Counters on the web services performance objects measure data transmitted at the OSI Application Layer.
- Counters on the TCP object measure data transmitted at the Transport Layer.
- Counters on the IP object measure data at the Network Layer.
- Counters on the Network Interface object measure data at the Data Link Layer.

For example, the counters at the Application Layer count the bytes sent before the data is divided into packets and prefixed with protocol headers and control packets because the data is in this form when the application sends it. Counters at the Application Layer do not include retransmitted data.

In addition, the counters display the data in units native to the component measured. For example, the Web Service object displays data in bytes, and the TCP object displays data in segments.

Optimizing Server Performance — Normal Load

During performance testing, HP used the following procedures to improve web server performance.

> **Caution**
> Some of the following modifications require the use of the Registry Editor. Using the Registry Editor incorrectly can cause serious, system-wide problems that may require you to reinstall Windows 2000 to correct them. Use the Registry Editor and these modifications at your own risk. Modifying the registry might require administrator access rights.

- **Remove any nonessential network protocols** — IPX\SPX and NetBEUI protocols can be removed to minimize system overhead.
- **Remove any nonessential system services** — Plug and Play, Alerter, and ClipBook Server services can be disabled to minimize system overhead.
- **Optimize performance for applications** — Set the *Application* setting, instead of the default *Background Services* setting in the System Performance Options. This instructs Windows 2000 to give more processor resources to the foreground program than to the background program.
- **Increase TCP/IP user port and window size registry value** — These values denote the number of user ports that TCP/IP can use to send and receive data, as well as the window size through which TCP/IP sends data.
  - `Hkey_local_machine\System\CurrentControlSet\Services\TCPIP\Parameters`
  - Add the parameters:
    - MaxUserPort: (DWORD)0xfffe (hex)
    - TCPWindowSize: (DWORD) 0x4470 (hex)
- **Release processor control on specific network cards registry value** — This modification parameter releases network card control from any particular processor unit and enables NIC interrupts and DPCs to remain on a single processor. As a result, this modification can decrease the interprocessing-interrupt calls between processors, especially within multiprocessing systems.
  - `Hkey_local_machine\System\CurrentControlSet\Services\NDIS\Parameters`
  - Modify the parameter ProcessorAffinityMask: (DWORD) 0x0000 (hex)

Optimizing Server Performance — High-Volume Websites

During performance testing, HP used the following procedures to improve web server performance for high-volume web servers. These modifications should only be used for high-volume websites. All registry modifications might not apply to your specific system.

- **Configure IIS for more than 100,000 hits/day** — This setting increases the amount of system resources available to the IIS for website use.

  This setting is located in the Internet Services Manager of the default website of the web server. Select *Properties* → *Performance* → *Performance tuning* increase the setting to *More than 100,000* expected hits per day.

- **Increase IIS ListenBackLog registry value** — This value denotes the number of connection requests in the queue for each service.

  - `Hkey_local_machine\System\CurrentControlSet\Services\Inetinfo\Parameters`
  - Add the parameter: ListenBackLog: (DWORD) 1000 (decimal)

- **Increase IIS AcceptExOutstanding registry value** — This value reflects the amount of AcceptEx sockets that IIS can use for incoming requests. This modification can decrease the amount of time incoming requests must wait in the request queue before being serviced.

  - `Hkey_local_machine\System\CurrentControlSet\Services\Inetinfo\Parameters`
  - Add the parameter: AcceptExOutstanding: (DWORD) 200 (decimal)

- **Increase IIS OpenFileInCache registry value** — This value denotes the amount of file handles that IIS can use in memory, instead of from disk. Such modification decreases overhead I/O calls to the server disk subsystem.

  - `Hkey_local_machine\System\CurrentControlSet\Services\Inetinfo\Parameters`
  - Add the parameter: OpenFileInCache: (DWORD) 1000 for every 32MB physical RAM (decimal)

- **Decrease IIS ObjectCacheTTL registry value** — This value denotes the amount of time that an inactive file stays in memory. This modification conserves the system resources for other IIS processes.

  - `Hkey_local_machine\System\CurrentControlSet\Services\Inetinfo\Parameters`
  - Add the parameter: ObjectCacheTTL: (DWORD) 180 (decimal)

- **Increase IIS MemoryCacheSize registry value** — This value denotes the amount of physical memory that IIS can use to cache system handles, directory listings, and binary large objects to improve performance.
  - `Hkey_local_machine\System\CurrentControlSet\Services\Inetinfo\Parameters`
  - Add the parameter: MemoryCacheSize: (DWORD) 1/2 of available physical memory in bytes (decimal)
- **Decrease IIS ProcessorThreadMax registry value** — This value denotes the number of threads/processor that IIS can allocate for Microsoft Transaction Server. Decreasing this value can decrease the number of thread contentions for processor time.
  - `Hkey_local_machine\System\CurrentControlSet\Services\W3SVC\ASP\Parameters`
  - Add the parameter: ProcessorThreadMax: (DWORD) 5 (decimal)
- **Increase IIS RequestQueueMax registry value** — This value reflects the number of ASP file requests that can be held inside the request queue for servicing. This modification decreases connection drops from the server platform when server request hits are high during the day.
  - `Hkey_local_machine\System\CurrentControlSet\Services\W3SVC\ASP\Parameters`
  - Add the parameter: RequestQueueMax: (DWORD) 1000 (decimal)
- **Enable IIS ASP buffering** — This modification allows the server to deliver the entire contents of a requested ASP file in one large packet-signal to the client platform as opposed to sending them in small chunks as the server finishes generating them.
  - `Hkey_local_machine\System\CurrentControlSet\Services\W3SVC\ASP\Parameters`
  - Add the parameter: BufferingOn: (DWORD) 1 (decimal)
- **Increase IIS ScriptEngineCacheMax registry value** — This value denotes the maximum number of ActiveX language engines that ASP will keep cached in memory. This modification increase ASP memory cache space and server ASP performance.
  - `Hkey_local_machine\System\CurrentControlSet\Services\W3SVC\ASP\Parameters`
  - Add the parameter: ScriptEngineCacheMax: (DWORD) 100 (decimal)

- **Increase IIS ScriptFileCacheTTL registry value** — This value denotes the amount of time in seconds that an inactive ASP script file will remain cached in memory before being deleted. This modification decreases disk I/O calls to disk for highly requested ASP files on the server platform.
  - `Hkey_local_machine\System\CurrentControlSet\Services\W3SVC\ASP\Parameters`
  - Add the parameter: ScriptFileCacheTTL: (DWORD) 1800 (decimal)
- **Decrease IIS SessionTimeout registry value** — This value denotes the amount of time in seconds that a session object is maintained in memory after the last request associated with the object is made. This modification frees system resources quicker for other IIS process usage.
  - `Hkey_local_machine\System\CurrentControlSet\Services\W3SVC\ASP\Parameters`
  - Add the parameter: SessionTimeout: (DWORD) 300 (decimal)
- **Increase IIS ThreadCreationThreshold registry value** — This value denotes the number of ASP requests that can be maintained in the common queue before a new thread is created in the thread pool. This modification decreases processor calls and overhead within the server platform.
  - `Hkey_local_machine\System\CurrentControlSet\Services\W3SVC\ASP\Parameters`
  - Add the parameter: ThreadCreationThreshold: (DWORD) 20 (decimal)

Learning Check

1. What are the four main factors that affect web server performance?

   .....

   .....

   .....

   .....

2. What are the three main capacity planning metrics?

   .....

   .....

   .....

3. What are the HP recommendations for capacity planning in relation to sizing memory?

   .....

   .....

4. It seems that a lengthy processor queue forms when servicing a large number of connections. What can you do ensure that accepted connections are processed promptly?

   .....

5. What System Monitor counter can be used to monitor the page file?

   .....

6. List three memory optimization procedures that you think will improve the performance of your web server.

   .....

   .....

   .....

Implementing Load Balancing

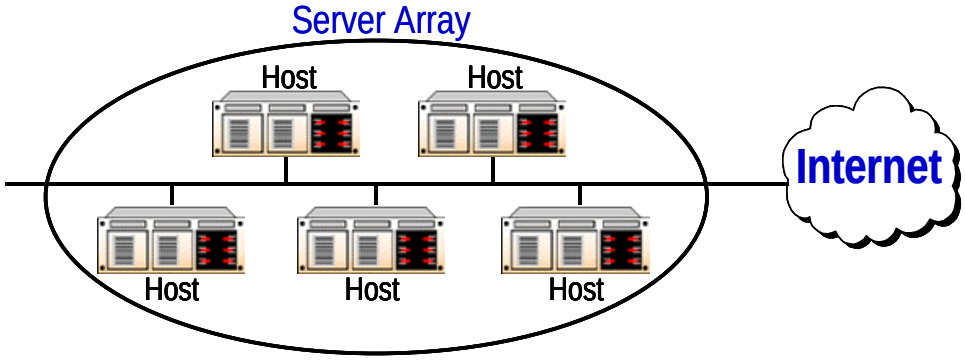
Module 6

Objectives

After completing this module, you should be able to:

- Describe the different load balancing methods.
- Describe HP network load balancing and availability solutions.
- List the key features of Microsoft Network Load Balancing (NLB) and explain how NLB works.
- Differentiate single NIC and multiple NIC NLB configurations.
- Explain how NLB servers exchange messages within a server array.

Load Balancing Overview



IP load balancing presents a single system image for clients in the form of a virtual IP address and distributes client request across multiple application servers.

Intelligent IP load balancing provides two advantages for the deployment of mission-critical Internet applications:

- **High availability** — The load balancer automatically detects failure of a server in the array and immediately routes traffic to another server. Availability is also ensured by the use of failover technology, such as a backup load balancer that takes over when the primary unit has failed. The failover capability prevents the load balancer from becoming a single point of failure.
- **Scalability** — Scaling an application can be as simple as plugging a pre-configured server into the network and allowing the IP load balancer to distribute the work load. In more complex applications, intelligent IP load balancers can balance load across multiple application servers that might have different processing power and I/O capabilities. This allows administrators to mix servers of various capabilities in the server array without the risk of overloading servers with smaller processing capabilities.

HP Load Balancing and Availability Solutions

Teaming NICs allows you to load balance the network traffic between the NICs. Additional network cards do not improve performance if the current NIC is not close to saturation.

Before you configure teaming NICs, consider the following:

- NICs must be alike, using the same driver and bus interface, and must be on the same network.
- In a load-balancing configuration, NICs must run at the same speed.
- NICs must use proper protocol configuration:
  - Do not depend on a DHCP server to assign an IP address.
  - Do not depend on a static NIC configuration to ensure proper network function.

HP ProLiant Ethernet network adapters support the following three types of teaming:

- Network Fault Tolerance (NFT)
- Transmit Load Balancing (TLB)
- Switch-Assisted Load Balancing (SLB)

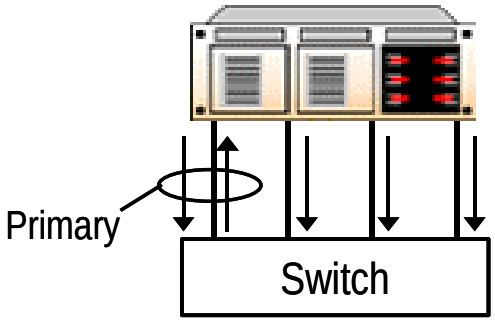
The way in which ProLiant network adapters are teamed depends on the team type. Most NC-series ProLiant Ethernet network adapters can be teamed with any other NC-series ProLiant network adapter.

Network Fault Tolerance

NFT is a simple, effective, and fail-safe approach to increase the reliability of server connections. NFT allows you to set up link recovery to the server NIC in case of a cable, port, or NIC failure. Each server can support up to eight teams where one adapter per team is defined as the primary adapter. All other adapters are secondary. By assigning at least two HP NICs as a team, NFT allows you to maintain uninterrupted network performance.

NFT teaming does not require ProLiant network adapters to share features. Any media and any speed can be used. For example, common NFT Teaming consists of a ProLiant Gigabit Ethernet adapter paired with an embedded 10/100 Mb/s Ethernet network adapter. The Gigabit Ethernet adapter is used as the primary adapter for greater bandwidth and performance.

Transmit Load Balancing



TLB uses up to four HP Intel-based Fast Ethernet NICs configured as a TLB team in the server. One NIC is primary and the other NICs are secondary. The primary NIC transmits and receives data like a typical Fast Ethernet NIC. The secondary NICs only transmit data. The TLB agent adjusts the data flow evenly between the multiple NICs. An algorithm that uses the last 1 or 2 bits of the source and destination memory access controller (MAC) or IP addresses determines which port is used for a particular server-to-client communication flow.

Example

A TLB Team containing four HP Fast Ethernet NICs, configured for full-duplex operation, can provide an aggregate, maximum transmit rate of 400Mb/s and a 100Mb/s receive rate. In this example, total bandwidth is 500Mb/s.

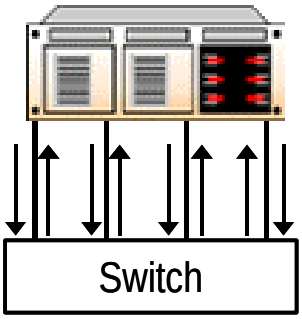
All NICs in the team operate under the same IP address, but different MAC addresses. If one of the secondary connections fails, the server driver redirects the information flow from the failed connection to the remaining NIC team members. If the primary NIC in the team fails, one of the secondary NICs assumes the MAC address and the duties of the primary NIC.

With TLB, traffic received by the server is not load balanced. The primary adapter is responsible for receiving all traffic destined for the server.

Note

Transmit Load Balancing is not supported with HP Netelligent or NetFlex-3 NICs.

Switch-assisted Load Balancing



SLB incorporates all the features of NFT and TLB. It supports up to four Fast Ethernet NICs in the server and balances the data load between NICs. If any of the NICs fail, the remaining NICs share the data load.

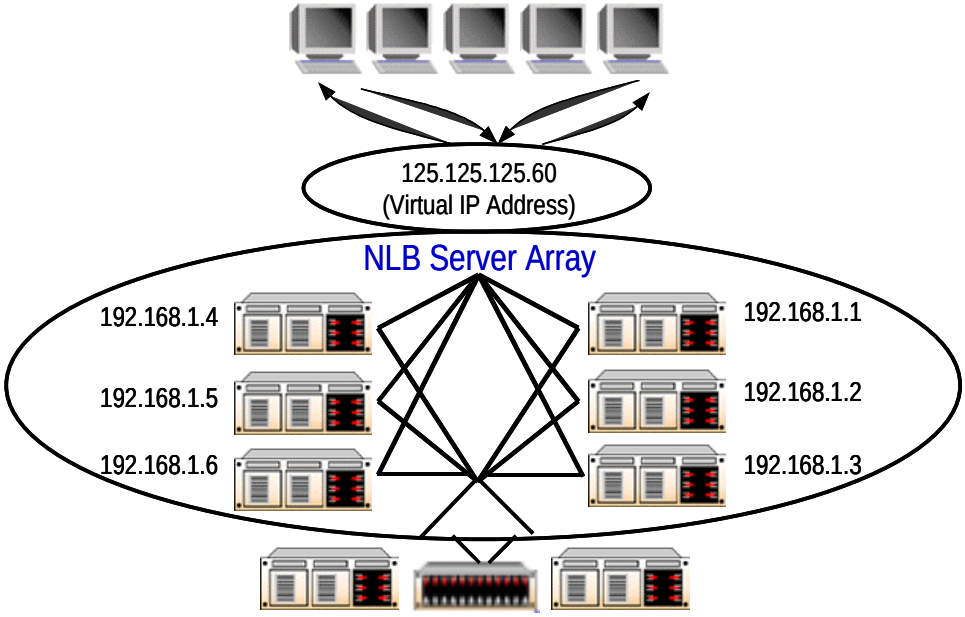
With SLB Teaming, there are from two to eight network adapters in a team and up to eight teams in a server. The operating system sees the multiple NICs as the same IP and MAC address, and, therefore, as one NIC. All adapters transmit and receive data simultaneously using the same speed, allowing for a transfer rate of up to 800 Mb/s. This approach must be used in conjunction with an intelligent switch that supports this type of teaming, and all ports must be connected to the same switch.

Example

Most switches require that all team members have identical capabilities. However, in an SLB or TLB configuration you can team a copper Gigabit adapter (operating at 1000 Mb/s) with a 10/100 Mbps Ethernet adapter, because both adapters are capable of 100 Mb/s, even though they may not be operating at 100 Mb/s.

On the other hand, teaming a fiber gigabit Ethernet network adapter and a 10/100 Mb/s Ethernet network adapter is not permitted because the fiber Gigabit adapter is not capable of operating at any speed at which the 10/100 Mb/s Ethernet adapter operates.

Microsoft Windows Network Load Balancing



NLB is integrated in Microsoft Windows 2000 Advanced Server, Microsoft Windows 2000 Datacenter, and Microsoft Application Center 2000. However, it must be installed on Microsoft Windows 2000 Server through Application Center.

NLB combines the resources of two or more Windows 2000 servers into a single front-end server array. NLB is a solution for web servers, FTP servers, HTTP proxy servers, and other IP-based services.

Each host in a web server array runs a separate copy of a web application such as Commerce Server or Content Management Server. These applications can use independent copies of databases residing within their individual servers, or they can share access to a separate, networked database server (such as in the DISA configuration). NLB distributes the work load among each host.

Clients access the server array as if it is a single computer by using one IP address. Under normal operations, NLB automatically balances the networking connections between the servers in the array. However, you can manually configure load distribution by modifying the NLB parameters.

When a computer fails or goes offline, NLB automatically reconfigures the array to direct the client connections to the remaining computers. The offline computer can transparently rejoin the array and regain its share of the work load at a later time.

Note

In the NLB documentation, the server array is referred to as a “cluster.” To avoid confusion with HP clustering solutions, the term “server array” is used in this course.

How NLB Works

NLB requires the following:

- Two or more servers with valid Microsoft Windows 2000 licenses
- Windows TCP/IP protocol installed
- One or more NICs in each server
- Ethernet or FDDI-based LAN

NLB uses less than 1MB of storage space and uses between 250K and 4MB of RAM during operations, assuming normal network load. The parameters can be modified to allow up to 15MB of RAM.

NLB installs as a standard Windows 2000 networking device driver (NDIS). When installed, it operates in a transparent manner to server applications and TCP/IP clients.

> **Note**
> When multiple network adapters are present, NLB should be installed on only one adapter.

NLB employs a fully distributed algorithm to statistically map incoming clients to the array servers based on their IP address, port, and other information. When inspecting an arriving packet, all servers simultaneously perform this mapping to quickly determine which server should handle the packet. The mapping remains invariant unless the number of servers changes.

The NLB filtering algorithm is efficient in its packet handling. Incoming packets are never modified and then retransmitted. On each server in the array, the NLB driver acts as a filter between the NIC driver of the array and the TCP/IP protocol stack to allow a portion of the incoming network traffic to be received by the local host. This feature allows the NLB to deliver high availability and scalability.

If you disable multicast support (causing the host to revert to unicast mode), NLB automatically instructs the driver belonging to the server array adapter to override the unique, built-in network address of the adapter and to change its MAC address to the MAC address of the server array, which is the address used on all hosts.

Some network adapters do not support changing their MAC addresses. If you experience this problem, you must install a network adapter that does.

Availability

NLB delivers high availability by redirecting incoming network traffic to working servers as necessary after a server in the array fails. Outstanding connections to a failed server are lost, but the network service remains available. In most cases (with web servers, for example), client software automatically retries the failed connections, and the clients experience only a few seconds delay in receiving a response.

Traffic Distribution and Control

NLB controls the distribution and partitioning of TCP and UDP traffic from Internet clients to selected servers within an array. After NLB has been configured, all servers within the array receive incoming client requests addressed to the IP address of the array.

NLB filters incoming packets to specified TCP and UDP ports before these packets reach the TCP/IP protocol software. NLB manages only the TCP and UDP protocols within TCP/IP, and its actions are controlled on a per-port basis.

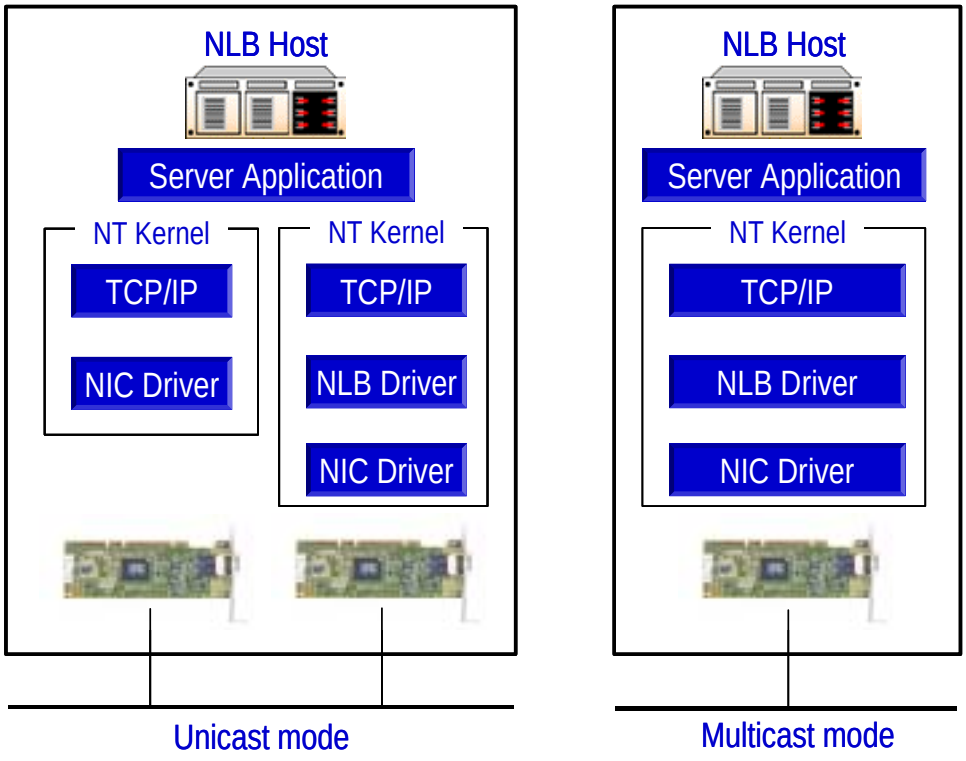
Scaling

NLB delivers scaled performance by redistributing the incoming network traffic among multiple servers. These servers respond to client requests based on *affinity*, the method used to associate requests from specific clients to specific servers. Affinity is configurable through the NLB Setup dialog box that is accessed through the Control Panel on each server.

When no affinity is specified, all network requests are load balanced across the server array without respect to their source. This maximizes the scaled performance. Affinity is accomplished by directing all client requests from the same IP address to the same array server.

Affinity is useful when load-balancing applications require the session state to be maintained. An example is an application that uses the Secure Socket Layer (SSL) protocol to provide encrypted communication. Affinity allows the negotiation of security information to occur each time a new host is encountered.

NLB Configurations



For maximum network performance, NLB typically operates in a dual-NIC configuration where a dedicated NIC handles array traffic while the other NIC handles other network traffic to the server. This type of configuration is referred to as unicast mode. NLB can operate in unicast or multicast mode.

! Important

NLB does not support a mixed unicast/multicast environment. If the host configurations are not consistent, NLB will not function properly.

In multicast mode, the MAC address of the server array is assigned to the network adapter that is functioning as the server array adapter. However, the built-in address of the server array adapter is retained so that both addresses are used. The first MAC address is used for client-to-server array traffic and the second for network traffic specific to the computer.

Multicast support is not enabled by default. However, if you do not enable multicast support, consider using at least two network adapters with one network adapter dedicated to handling client-to-server array traffic.

Example

Devices on the network are managed by HP Insight Manager 7, which discovers devices using web-enabled agents or pings. If an NLB host has a single NIC, HP Insight Manager discovery, polling, and software deployment meant for an individual server could interfere with all hosts in the server array if multicast support is not enabled.

NLB can also operate in a multi-homed configuration in which a host has multiple NICs that have individual IP addresses or a single NIC with multiple IP addresses. This enables NLB to load balance multiple applications running on the same servers.

Port Rules

Port rules allow you to control how traffic is filtered. Each host in the server array must have the same port rules. A host cannot join a server array that does not have the same port rules.

Network traffic can be handled by multiple hosts, a single host, or disabled for a specific port rule. When multiple hosts handle network traffic, you can specify that the load be equally distributed among the hosts or that each host handles a set load.

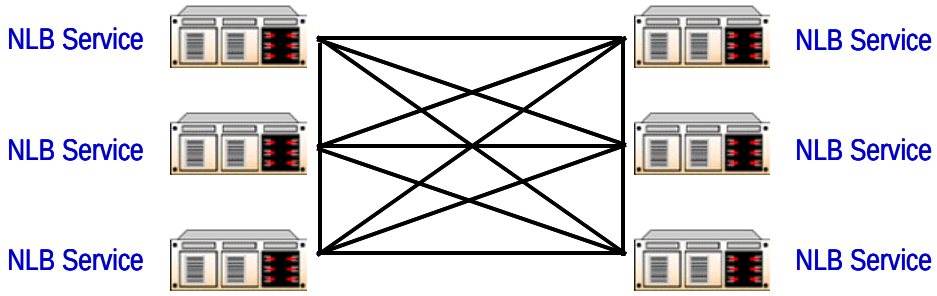
Example

A web browser obtains different images for a single web page from different servers within an NLB array. This process distributes the server load, speeds up client-response processing, and shortens the overall response time to clients.

One computer handles all network traffic when the filtering mode is set to *single host*. The computer with the highest priority handles all the traffic for the associated port rule.

To block unwanted network access to a range of ports, disable network traffic for a specific port rule. The NLB driver filters all corresponding network packets and datagrams associated with that port rule.

Coordination Within an NLB Array



To coordinate their actions, NLB servers periodically exchange multicast or broadcast messages within the array. This allows the servers to monitor the status of the array.

NLB assumes that a server in the array is functioning properly as long as it participates in the normal message exchange among the other servers. When the state of the array changes (such as when servers fail, leave, or join the array), NLB invokes a process known as *convergence*. If other servers do not hear from any member for several periods of message exchange, convergence is initiated to redistribute the load previously handled by the failed server.

The system administrator can control both the message exchange period and the number of missed messages required to initiate convergence. The default values are respectively set to 1,000 milliseconds (1 second) and 5 missed messages. Because these parameters are not usually modified, they are not included in the normal NLB Setup dialog box. The parameters are adjusted manually in the registry as necessary.

The keys to adjust these parameters are:

```
HKEY_LOCAL_MACHINE\System\CurrentControlSet\Services\WLBS
    \Parameters\AliveMsgPeriod
```

and

```
HKEY_LOCAL_MACHINE\System\CurrentControlSet\Services\WLBS\Parameters
    \AliveMsgTolerance
```

Convergence

Convergence is initiated only when a server fails and not when a service or port fails. During the process of convergence, NLB servers exchange messages to determine a new, consistent state of the server array and to elect the server with the highest server priority, known as the *default server*. When all host servers have reached consensus on the correct new state of the array, they announce the completion of convergence in the Windows 2000 event log.

During convergence, the servers continue to handle incoming network traffic as usual, except that traffic for a failed server does not receive service. Client requests to working servers are unaffected. At the completion of convergence, the traffic for a failed server is redistributed to the remaining servers. Load balanced traffic is repartitioned among the remaining servers to achieve the best possible new load balance for specific TCP or UDP ports.

If a server is added to the array, convergence allows this server to take over ports for which it has the highest priority and to receive its share of the load-balanced traffic. Expansion of the array does not affect ongoing host operations and is achieved transparently to both Internet clients and server applications.

Learning Check

1. What are the three types of commercial IP load balancing products?

   .....

   .....

   .....

2. Into which of these three categories does NLB fit?

   .....

3. What process is initiated when the state of the server array changes?

   .....

4. What method does NLB use to associate requests from specific clients to specific servers?

   .....

5. How do NLB servers coordinate their actions?

   .....

   .....

Objectives

After completing this module, you should be able to:

- List the requirements, roles, and responsibilities of an information security policy.
- Identify vulnerabilities, common attacks, and measures to counter associated attacks.
- Describe the features of ISS Internet Scanner.
- Describe the best practices for securing a web server in regard to:
  - Physical security.
  - Network security.
  - Host security.
  - Application security.
  - Website data security.

Information Security Policies

The purpose of an enterprise Information Security Policy is to define the appropriate measures that must be employed to manage the operational risk associated with the information systems of a company. Risk management is a complex issue involving a variety of potential activities. In this policy, risk management is focused on risks arising from generic internal business activities of an enterprise and the risks associated with external network interfaces, related to the operation of an enterprise-wide information system.

A security policy does not create a secure environment. The policy must be properly implemented using relevant procedures and technology to have an impact on risk.

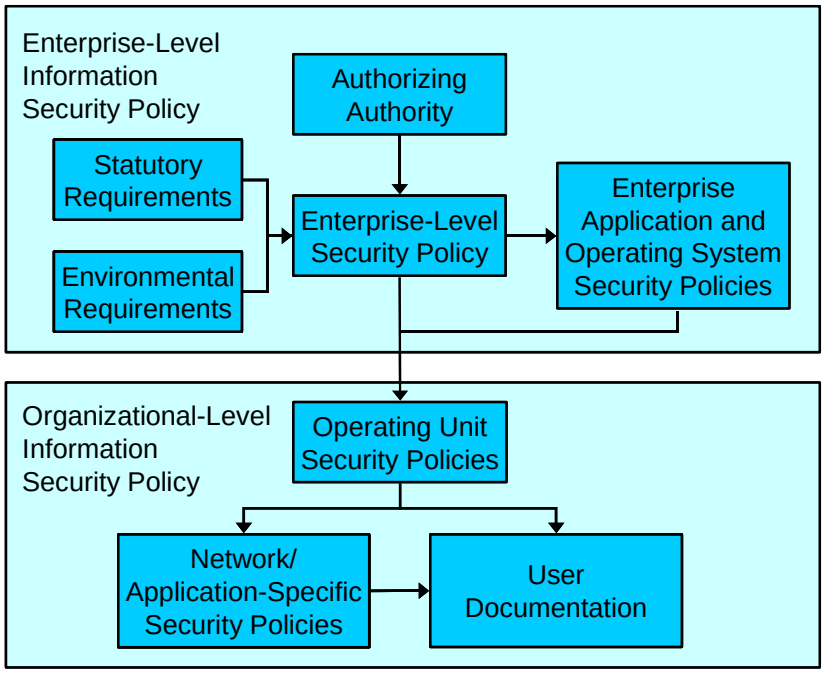
This enterprise Information Security Policy focuses on the following goals:

- To provide a level of system availability that is sufficient to ensure appropriate support for performance of the enterprise mission.
- To provide appropriate protection of the ownership and confidentiality of sensitive information, including proprietary, personal, and confidential information.
- To provide appropriate data integrity for information used in critical business decisions.

In addition, tracking system usage using audit trails can be used to determine the source of problems that may arise. These records can also be used for billing, capacity planning, and other purposes.

Hierarchy of Information Security Policies

A hierarchy of information security policies usually exists within the security profile of an enterprise that supports each level. Policies begin at a very high level and become increasingly detailed as they apply to operational environments.



An authorizing authority document is issued from a company executive to establish an information security program and to delegate the authority to an information security program manager who coordinates and develops an enterprise-level security policy that applies to the entire company. The information security program manager may also issue security policies for enterprise applications and operating systems.

Enterprise-level security policies should consider the operational environment and any applicable statutory requirements imposed by federal, state, and local governments. To support development and implementation of security policies, a security policy board may be set up to give organizations within the enterprise a voice in policy development and to build a consensus to support implementation.

Enterprise-level policies are usually applied at the organization level. When policies apply to specific hardware or software components, the policy descriptions can be very detailed and provide explicit guidance, which is implemented through user documentation. The network-specific policy lists products and provides instructions on how they are to be set up and maintained.

The enterprise-level and the organizational-level policy are often the same document in small and medium-size companies. Each enterprise molds its policy model to fit, and as the enterprise grows, the model evolves. Therefore, flexibility is key and should be considered when developing and implementing an information security policy, which is the first step in security planning.

Security Vulnerabilities

The majority of successful attacks on computer systems by means of the Internet can be attributed to exploitation of security flaws. Hackers are opportunistic, taking the easiest and most convenient route. They exploit the best-known flaws with the most effective and widely available attack tools. They count on organizations not fixing the problems, and they often attack indiscriminately, scanning the Internet for any vulnerable systems.

System administrators typically do not correct many system flaws because they do not know which vulnerabilities are most dangerous and are too busy to correct them all. Some vulnerability scanners search for up to 800 vulnerabilities, which broadens the focus needed to ensure that all systems are protected against the most common attacks.

The SANS Institute and the National Infrastructure Protection Center (NIPC) released a document summarizing the Most Critical Internet Security Vulnerabilities. The following software vulnerabilities account for the majority of successful attacks:

- Default installs of operating systems and applications
- Accounts with no passwords or weak passwords
- Nonexistent or incomplete backups
- Large numbers of open ports
- Not filtering packets for correct incoming and outgoing addresses
- Nonexistent or incomplete logging
- Vulnerable Common Gateway Interface (CGI) programs

The ability to scan a computer for open IP ports is an unavoidable vulnerability because of the idiosyncrasies of the IP protocol. Additionally, unknown vulnerabilities, such as a buffer overflow attack on a DNS server, exist that must be predicted. If an anomaly exists in the DNS server, sending a command with a long hostname could cause the DNS server to fail.

Common Attacks

Following is a list of common attacks used by *hackers*:

- **Dictionary** — A program used to test a security accounts database against a long list of names or passwords. These attacks rely on automated programs.
- **Man-in-the-middle** — A hacker detects passwords and obtains information from legitimate transactions.
- **Hijacking** — A hacker trespasses in a transaction between two parties, imitates one of the participants, then continues the connection.
- **Trap door misuse** — A piece of residual code in the system that allows the original programmer access to easily support the product. However, trap doors make the system vulnerable to hacker exploitation.
- **Back door** — An undocumented opening in an operating system or program that is programmed with benign or malicious intentions.
- **Denial-of-service (DoS) attack** — A hacker makes it impossible for a legitimate user to access or service a host.
- **Buffer overflow** — If a user sends more data than the target system can receive at one time, the server will fail. The extra data overflows the storage buffer of a program and then overwrites the actual program data. This action allows the program of a target system to be modified spontaneously and remotely.
- ***Trojans* and *worms*** — Files that operate in an expected way, but also have a secret operation that subverts security. Trojans include files that can send sensitive information back to a hacker. Worms propagate from system to system.
- ***Virus*** — Programming code that, when run, attaches to other programs and executes each time the infected files run.
- **Exploiting default settings** — The default settings of software before it has been configured.
- **IP spoofing** — An attempt to defeat authentication. It requires a packet *sniffer* and a program to kill a TCP connection, generate a TCP connection, and spoof (mimic) an IP packet.
- **Blind and nonblind spoofing** — Blind spoofing occurs when a hacker manipulates a connection that exists on a separate physical line. Nonblind spoofing occurs when a hacker manipulates a connection on the same subnet or the same physical line.

- **SYN flood** — A hacker creates multiple half-open TCP connections. When the SYN queue is flooded, no new connection can be opened.
- **Smurf attack** — A smurf attack involves manipulating ICMP, the protocol invoked by the ping program. It is a form of IP spoofing that results in a DoS attack.
- **Teardrop** — A teardrop attack takes advantage of code that does not properly reassemble overlapping UDP packets.
- **Windows out-of-band** — An out-of-band DoS attack is used to cause the computer to fail or a loss of network connectivity.
- **Land attack** — A hacker sends a TCP SYN packet with a spoofed (fake) source IP address and port number that matches that of the destination IP address and port to attempt to cause the networking to loop and eventually cause the computer to fail.
- **Ping of death** — A large amount of information is appended to an Internet Control Message Protocol (ICMP) echo request (ping) packet to cause a kernel buffer overflow when the computer attempts to respond, which causes it to fail.
- **Port scan** — A port scan is used to access the open IP ports on a target computer or network.
- **IP half scan** — An IP half scan is used to avoid detection. TCP communication uses a three-step process to establish a connection. The IP half scan completes only half the connection, avoiding detection.
- **UDP bomb** — A hacker sends an illegal User Datagram Protocol (UDP) packet. The UDP packet contains incorrect values that may cause older operating systems to fail.
- **Website defacement** — A hacker defaces a website by changing or destroying the content.

  One of the most notable website defacements was an attack of the website of the U.S. Army Research Laboratory. The website was altered with a message criticizing the military organization for supplying weapons to Israel.
- **Cyber-extortion** — Cyber-extortion involves hackers blackmailing companies by threatening to turn over purloined strategic data to their competitors or expose it to the public. The more popular cases resembled kidnapping in which data, such as customer personal information and credit card numbers, were held hostage in exchange for money.

Potential Security Vulnerabilities in SNMP

Simple Network Management Protocol (SNMP) is an industry-standard protocol defined by the Internet Engineering Task Force.

The Computer Emergency Response Team (CERT) at Carnegie Mellon University published a Security Advisory regarding a SNMP vulnerabilities.

INTERNET The CERT article can be found at:

http://www.cert.org/advisories/CA-2002-03.html

CERT outlined vulnerabilities that can cause SNMP services to stop functioning and in some cases may enable unauthorized access, DoS attacks, or may cause system instability.

The SNMP software causing part of this problem is delivered as a component of the operating system. Therefore, the nature of the fix is in the form of an operating system update or patch.

HP

HP uses the industry-standard SNMP layer software as a communications protocol between the HP Management Agents and HP Insight Manager 7. HP Management Agents also use SNMP to communicate with third-party management products.

If a system managed by Insight Manager is attacked, Insight Manager will not be able to communicate with the system to get status information and the system status will change to “Unknown”. In addition, the affected system will not be able to send alerts (known as SNMP traps) to the management system. Disabling SNMP will prevent the HP Management Agents from sending SNMP traps (alerts) and responding to SNMP status requests.

The primary functionality of the Remote Insight Lights-Out Edition (graphical remote console, virtual floppy, and virtual power) is not affected by the SNMP vulnerability. Only the ability to send SNMP traps and forward SNMP traffic from the HP Management Agents running on an attacked system through the Remote Insight Lights-Out Edition and Integrated Lights-Out are impacted.

Recommendation

Although SNMP is the most widely used protocol for management, it is not a very secure protocol. HP recommends not exposing unprotected computers running SNMP outside of your corporate or private network. Firewalls protecting LANs should be checked to ensure they are configured to block the ports that SNMP uses (port 161 and 162).

The SNMP community strings provide a method to identify systems that are authorized to communicate over SNMP. HP recommends changing the default community strings from “public” for SNMP get operations and “private” (for SNMP set operations). These community strings should be set to unique values that are changed frequently.

You can easily change SNMP community strings on a group of Microsoft Windows devices periodically (by means of a schedulable task).

Microsoft

All network operating system versions of Windows provide an SNMP implementation that is neither installed nor running by default. A buffer overrun is present in all implementations.

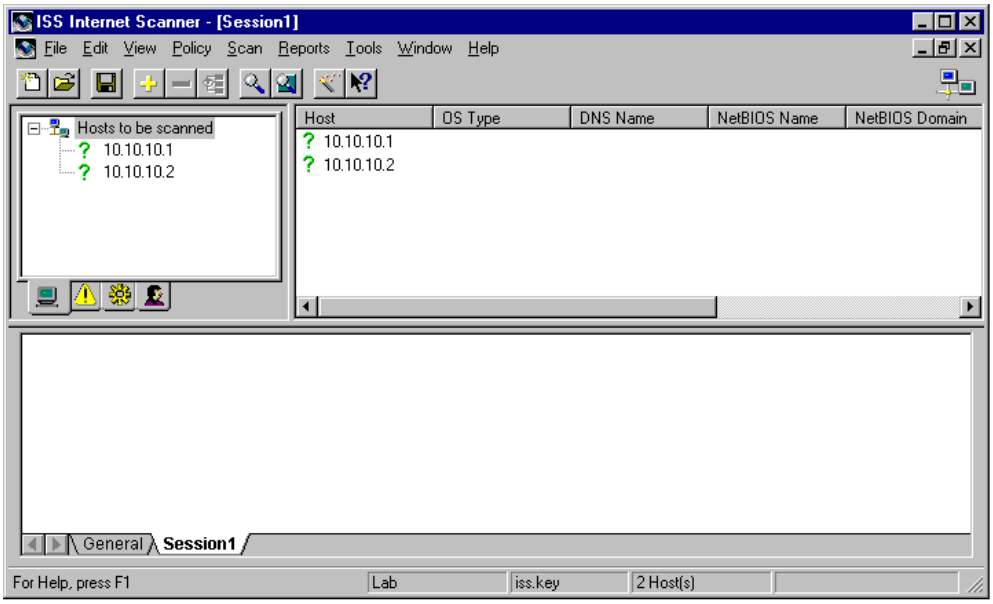
By sending a malformed management request to a system running an affected version of the SNMP service, an attacker could cause a denial of service. In addition, the code could run on the system in LocalSystem context, which could enable the attacker to take any desired action on the system.

The SNMP Parameters key, `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\SNMP\Parameters`, provides the SNMP community name and SNMP management station identifiers. SNMP community strings may allow either read or read-write access to the SNMP service. If no read-write access strings exist, the user could only use this vulnerability to read information through SNMP that is normally available to local users. If read-write access strings exist, a malicious user could use this vulnerability to make changes to any system. The default permissions could allow a malicious user to monitor or reconfigure certain devices on a network.

Recommendation

Microsoft Service Pack 2 for Windows 2000 corrects the permissions on several registry values in Microsoft Windows 2000.

ISS Internet Scanner



ISS Internet Scanner is a vulnerability assessment application that analyzes the security of devices on an enterprise-wide network. Internet Scanner has more than 30 predefined reports that help you collect the information you need to make security policy decisions.

You can generate data for reports in several ways:

- Schedule Internet Scanner to scan network devices automatically
- Launch Internet Scanner from the command line or the user interface
- Generate a report from the command line or user interface

Depending on the reporting options you select, Internet Scanner provides information about each detected vulnerability, including its location, detailed description, and suggested corrective actions.

Securing a Web Server

- Physical Security
- Network Security
- Host Security
- Application Security
- Website Data Security

To understand what security measures should be taken to secure your web server, you should have a clear understanding of the environment surrounding your web server.

A security policy should be developed that defines which users are allowed access to specific data and network resources and when they can access them. All installed security rules should be a reflection of the security policy.

The primary security layers are:

- **Physical security** — The physical location of the server and associated environmental issues
- **Network security** — The logical locations of the server and other network entities (routers, firewalls, and so on) and the application-level protocols running over the network (DNS or SNMP, for example)
- **Host security** — The security of the host server operating system
- **Application security** — The security of the web server application
- **Website data security** — The security of the website data

Physical Security

The proliferation of web-based applications increases the possibility that a web server is not under central control. The web server must be protected in a manner consistent with its value to the organization. Considerations include:

- Servers should be physically isolated from normal office traffic and protection measures should limit the number of people who have physical access to the server.

  A web server residing in an unsecured area is vulnerable to both intentional and accidental damage from a variety of sources that include theft, power cycling, physical damage, configuration tampering, and disconnection from the network.

- Virtually all office environments comply with local building codes, which effectively mitigate all but the most catastrophic environmental risks.

- Depending on the location and organization, protective forces such as building guards might be appropriate to limit access to the area in which the web server is housed. In addition, common-sense measures such as fire extinguishers, door locks, and climate controls, should be used.

  A web server residing in an area without proper climate control is subject to excess heat, humidity, and cold that could damage the server.

- Surge protectors should be installed on all servers, and mission-critical servers may require uninterruptible power supplies (UPSs).

  A web server without an adequate contingency power supply could suffer physical damage because of sudden loss of power and subsequent power surges.

The goal of deploying a web server is to provide information and a method of interacting with real-time data for users not on your private network. Whether the cause is deliberate or accidental, anything that interferes with users' access to web data is considered a denial of service.

Recommendations

The following measures provide physical security for the server:

- Limit physical access to the web server.
- Locate the web server in an area that provides proper climate control.
- Connect a UPS to the web server to allow enough time to perform a proper system shutdown during a power interruption.
- Implement a high-availability solution using redundant web servers.

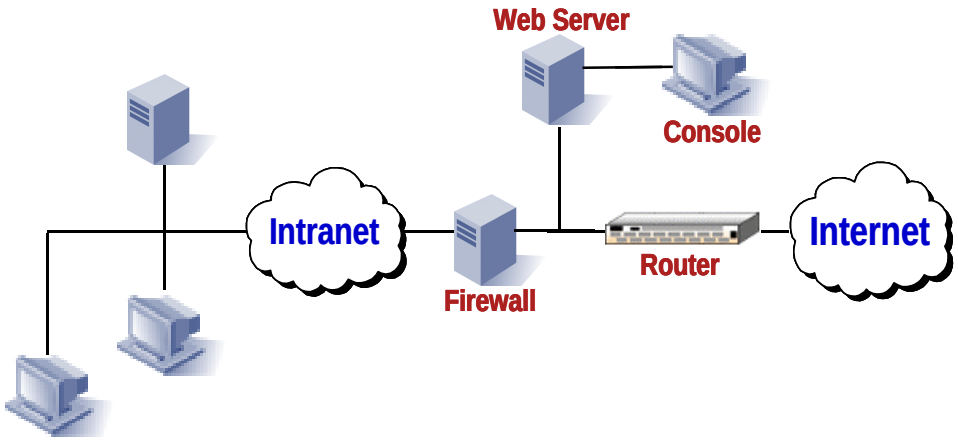
Network Security

The primary security issue in the network security layer is related to the location of the web server on the network. The way security technologies are implemented can significantly affect the scalability and availability of a DISA system. For example, placing a firewall between the application servers and the data resources can become a performance bottleneck depending on the security settings. The firewall itself is a potential single point of failure unless configured with an appropriate failover option.

In addition, not all firewalls support proprietary database protocols. For example, the application servers may not be able to access Oracle databases hidden behind certain firewalls.

The four traditional web server and firewall configurations are discussed on the following pages. Regardless of the configuration, it is critical that the firewall be properly configured. Otherwise, the web server is susceptible to various attacks such as SYN flooding or IP spoofing.

Configuration One — Web Server Outside the Firewall



The first traditional configuration places the web server on the public network between the firewall and the router connecting the network to the Internet. This configuration could also apply if there were no firewall in the network configuration.

The disadvantages of this configuration are:

- The web server is not protected by the firewall against:
  - Spoofing and protocol filtering
  - Content inspection
  - IP activity logging
  - Network address translation
- The web server is exposed to many types of network and protocol attacks such as:
  - SYN flood
  - Ping of death
  - IP spoofing
  - Denial of service
- Remote management and monitoring from inside the firewall is difficult and potentially insecure.

In this configuration, the remote management console that monitors the web server must be located outside the firewall because many firewalls do not support the protocols (such as SNMP) required for the monitoring operations. Therefore, the web server management console is vulnerable to the same threats as the web server.

The advantage of this configuration is that this is the best placement for web performance for high-volume sites.

With this configuration, the following recommendations apply:

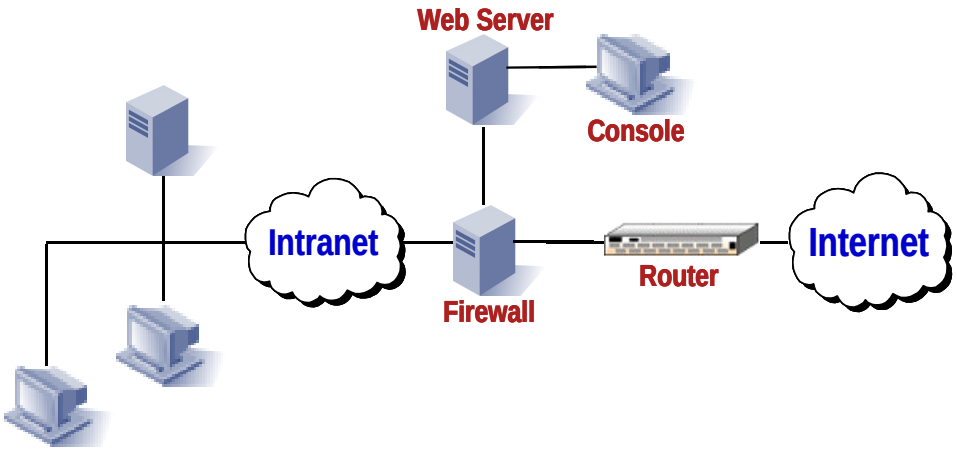
- Place the web server on a dedicated subnet and use the security features incorporated in the router to filter out all unnecessary protocols going to the web server subnet. For example:
  - Allow HTTP, FTP, and DNS traffic from the router interface to the web server subnet.
  - Allow all traffic from the web server subnet through the router interface.
  - Disallow all other traffic from the router interface to the web server subnet.
  - According to your security policy, allow traffic from the router interface to the public firewall interface.
  - According to your security policy, allow traffic from the firewall private interface through the router.
- If possible, connect the monitoring console directly to the web server.

These recommendations provide some measure of network security to the web server. However, they do not provide robust security because router security is not as robust as firewall security. The web server remains a potential target of a variety of IP attacks.

Some intelligent load balancers provide a base level of security protection for the rest of the DISA systems. Router-based load balancers block connections through all TCP/IP ports other than those the administrator chooses to open.

The router is also a target for attack. If router security is compromised, attackers can use sophisticated tools to probe for weaknesses in the web server. Best security practices should be applied to the router to prevent an attacker from compromising the security of the router. For example, you should limit the number of router administrators, avoid dual access passwords, and specify proper password lengths.

Configuration Option Two — Web Server Connected to the Firewall



In the second traditional configuration, the web server is connected to the firewall with a separate network interface.

The disadvantages of this configuration are:

- Performance might be degraded on high-volume sites because all traffic must now pass through the firewall and the firewall interface.
- The remote management and monitoring console for the web server must be located on the web server subnet or be connected directly to the web server because many firewalls do not support the protocols, such as SNMP, that are required to perform the monitoring operations.

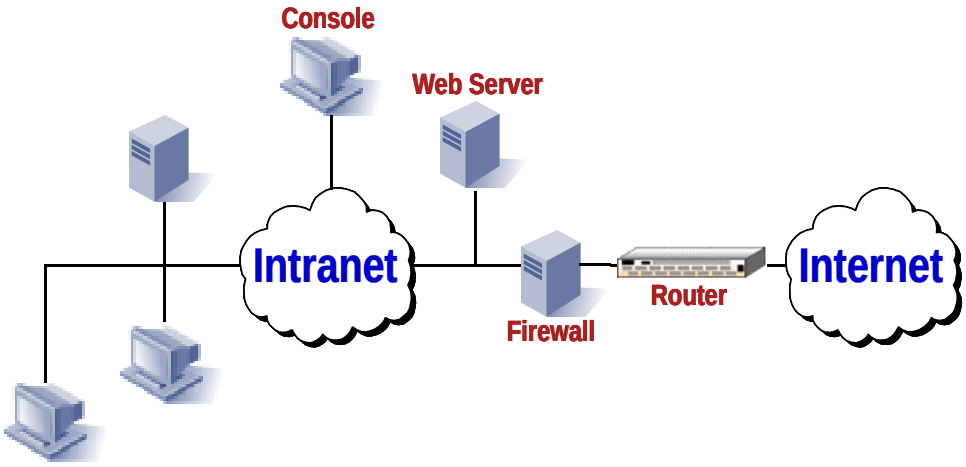
The advantages of this configuration include:

- The web server is protected against:
  - Spoofing
  - Protocol filtering
  - Content inspection
  - IP activity logging
  - network address translation
- The intranet is not at risk if the web server is compromised.
- The monitoring console is protected by a firewall.

Consider the following recommendations:

- Ensure that a firewall security policy has been defined as follows:
  - Filter out unnecessary protocols.
  - Configure the firewall to protect against IP spoofing.
  - Use content inspection where applicable.
  - Turn on logging and suspicious activity monitoring.
  - Use network address translation to protect the private network addressing.
- Periodically run a vulnerability assessment tool against the firewall.

Configuration Option Three — Web Server Located Behind the Firewall



In the third traditional configuration, the web server is located on the private network behind the firewall.

The disadvantages of this configuration are:

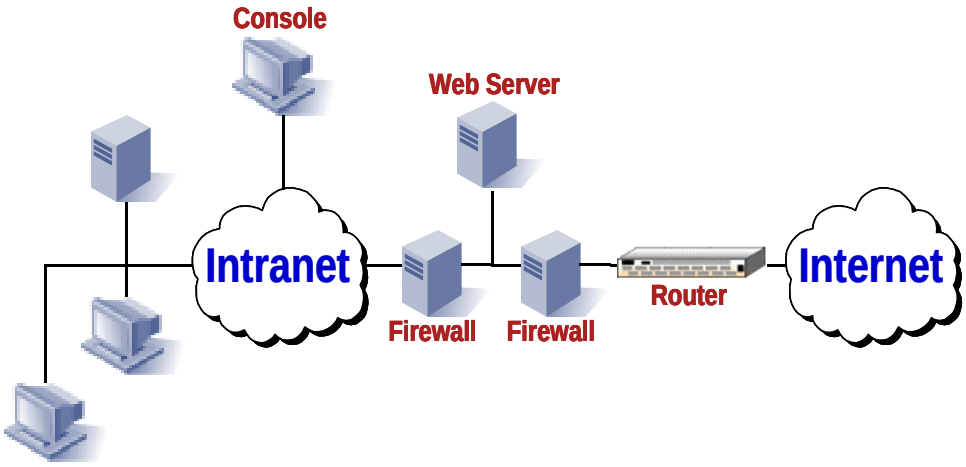
- The private network behind the firewall is at risk of attack because the web users are accessing a web server physically located on the private network. Remote attacks targeting the private network are possible.
- Performance might be degraded on high-volume sites because all inbound traffic must now pass through the firewall and the firewall interface.

The advantages of this configuration are:

- The web server is protected against:
  - Spoofing
  - Protocol filtering
  - Content inspection
  - IP activity logging
- The monitoring console can remotely monitor the web server from the relative safety of the private network.

> **Note**
> Although this option provides the web server some measure of security, it also exposes the private network to potential remote network attacks and is, therefore, dangerous and not recommended.

Configuration Option Four — Web Server Located Between Firewalls



In the fourth traditional configuration, the web server is located on a network segment between two firewalls.

The disadvantages of this configuration are:

- Remote management and monitoring is difficult. The remote management and monitoring console for the web server must be located on the same network segment as the web server or be connected directly to the web server because many firewalls do not support the protocols required to perform the monitoring operations.
- Configuration and administration are more complex, which can increase the chances of security exposure.
- This configuration is more expensive because it requires additional firewalls and supporting network devices.
- Relative performance might be degraded in this configuration because users must transit multiple firewalls to reach the Internet.

The advantages of this configuration are:

- The web server is protected by the firewall against common attacks.
- The internal network is not at risk if the web server is compromised.
- The internal network is still protected if one firewall is compromised.
- The monitoring console is protected by a firewall.

Consider the following recommendations:

- Ensure that a well-defined firewall security policy has been implemented.
- Periodically run a vulnerability assessment tool against the firewall.

Host Security

Host security refers to the security features built into the server operating system. Windows 2000 provides administrators with properly implemented security capabilities that:

- Isolate individual resources, especially data, by mediating access to those resources.
- Provide accountability of individual users to individual actions through audit trails and ensure the integrity of audit information.
- Isolate the system by enforcing identification and authentication of users, requiring the user to be positively authenticated through login procedures before any other activity is permitted.
- Prevent information left in memory and on storage peripherals from being scavenged after its legitimate usage has passed.

Discretionary Access Control (DAC) alone does not ensure the security or integrity of the Windows 2000 operating system. The web server must be properly configured with the appropriate Windows 2000 security features. Failure to do so will expose the operating system, and therefore the web server, to serious attack.

Recommendations

Configure Windows 2000-based web servers using the following guidelines:

- Format server hard drives for NTFS.
- Install the latest service pack and hotfixes
- Limit the number of Windows 2000 Server administrators.
- Create an account for each Windows 2000 administrator.
- Ensure that Windows 2000 administrators use encryption, token key devices, or similar technology for remote authentication.
- Log and audit all security events.
- Enable auditing for all sensitive operating system and web application files such as log files, password files, configuration files.
- Review audit trails on a regular basis.
- Perform backups using off-site storage.
- Disable the Windows 2000 Server service binding to any network adapter cards connected to the Internet.
- Review the default permissions of all Windows 2000 security groups. By default, the group Everyone, which contains all users and groups, including the Internet Guest account and the Guest group, has full control of all files created on the NTFS drive.
- Review directory and file permissions. Windows 2000 default directory and file permissions are typically too open to provide adequate security.
- Set all system files, script files, and data files to *no access* or *read only* for all users except system administrators. A special directory with write access can be set up for users if required.
- Require robust password controls (length, reuse, mixed character usage) on all accounts.
- Rename the Windows 2000 default Administrator account.
- Disable the Windows 2000 Guest account.
- Disable LAN Manager authentication and use *NTLM* version 2

In addition to the operating system-specific security recommendations, the following security issues should be addressed.

- Install antivirus software on the server with the real-time virus scanning option activated. Schedule batch virus scans to prevent virus infection from affecting the performance and availability of the server.
- Run IIS Lockdown tool.
- Write IIS logs to a dedicated partition

Application Security

Application security refers to the security implemented by the web server application software. IIS is fully integrated with and runs as a service of Windows 2000 Server. IIS takes advantage of the Windows 2000 security model and does not add security layers over Windows 2000.

IIS, which is required for all web applications, takes advantage of the Windows 2000 security model. However, options are available in IIS that can enhance security on the web server.

Poor implementation of security in the host security layer results in poor security in the application security layer. There can also be a tendency to ignore the security features in IIS because proper security has been implemented in the other three security layers.

Following are recommendations for implementing security at the application layer:

- Allow *execute only* permission on IIS virtual directories where users do not need read access.
- Allow *read only* permission on IIS virtual directories where users do not need execute access.
- Impose IP-based restrictions. IIS can restrict access from a specific IP address or addresses and allow access only from a specific IP address or addresses.
- Set the *Web Site Operators* option in the Default Web Site Properties tab to have only Administrators. Only administrators should be able to administer a website.

Recommendations for Sites Not Requiring Authentication

Consider the following recommendations:

- Enable the *Allow Only Anonymous Connections* option in the Default FTP Site Properties tab. This prevents users from logging on with user names and passwords that might include permissions that are not needed by the users.
- Set the *Enable Automatic Password Synchronization* option in the Default FTP Site Properties tab. This synchronizes the anonymous password and password settings with those set in Windows 2000.
- Review IIS web server logs on a regular basis.

Recommendations for Sites Requiring Authentication

Consider the following recommendations:

- Require authentication using Windows 2000 challenge/response, SSL client certificate mapping, or a custom third-party authentication using the ISAPI filter. With Windows 2000 challenge/response, the web server engages in a cryptographic exchange with the browser to verify the password sent.

  If SSL client certificate mapping is used, the X.509 client certificate provided by the user is verified and used as an index to look up a mapping to a Windows 2000 user account. The user does not need to enter a user name and password. This is a secure method of authentication as long as the X.509 certificate is issued from a trusted Certificate Authority.

  Other forms of authentication such as HTTP basic authentication, FTP authentication, and MCIS membership server verification send ID and password in clear text. This is not a recommended method for authenticating users.

- Control access to directories. Disable the *Directory Browsing Allowed* check box on the Directories tab of the Properties dialog box for the website.

Website Data Security

Protecting web data from unauthorized disclosure requires the data to be protected both on the server and in transit from the web server to the user's web browser.

The proper way to protect data on a server from unauthorized access is to require secure user authentication. To ensure confidentiality during transmission, the web data should be encrypted. Web server vendors are implementing various security protocols to support this requirement.

Recommendations

The web server administrator should require secure user authentication using HTTPS over a security protocol such as SSL. This will ensure that remote web users are properly authenticated with an ID and password that are encrypted and that the web content being transferred from the web server to the browser cannot be intercepted and read while in transit.

Note

The user's Internet browser must support the use of security protocols such as SSL for secure authentication and content encryption to occur. The most common browsers support the use of security protocols. SSL is the most widely used security protocol in the current web environment.

Anonymous login is not compatible with the concept of securing web data and should not be allowed on any web server attempting to secure the viewing of web content. While the access permissions of users using anonymous login can be restricted, no individual authentication or accountability is maintained. It also is not practical to encrypt data that is accessible by anonymous users because the data is easily compromised.

Learning Check

1. List three software vulnerabilities that account for the majority of attacks.

   .....

   .....

   .....

2. What role does ISS Internet Scanner play in the security policy of an enterprise?

   .....

   .....

   .....

3. List the primary security layers and briefly describe their functions.

   .....

   .....

   .....

   .....

   .....

   .....

4. What are the disadvantages of a web server outside of the firewall configuration?

   .....

   .....

   .....

   .....

5. What is host security?

   .....

   .....

6. What type of user authentication is required for a secure web server?

   .....

   .....

   .....

   .....

Objectives

After completing this module, you should be able to:

- Describe intrusion detection and explain the elements used in intrusion detection systems.
- Explain where intrusion detection servers should be placed in a network architecture.
- Describe the features of Internet Security Systems (ISS) RealSecure.
- Describe the RealSecure components and how sensors are deployed and managed.
- Explain how Microsoft Internet Security and Acceleration (ISA) Server detects and responds to intrusions.
- Describe the Malicious Activity Detection (MAD) feature of Check Point Firewall-1.

Intrusion Detection

Intrusion detection and vulnerability assessment are two key components of an enterprise security policy. To manage operational risk associated with a network, you need the ability to recognize and respond to an attack while it is happening. Otherwise, you cannot prevent the attacker from inflicting damage. An attack recognition and response system is required to address this issue.

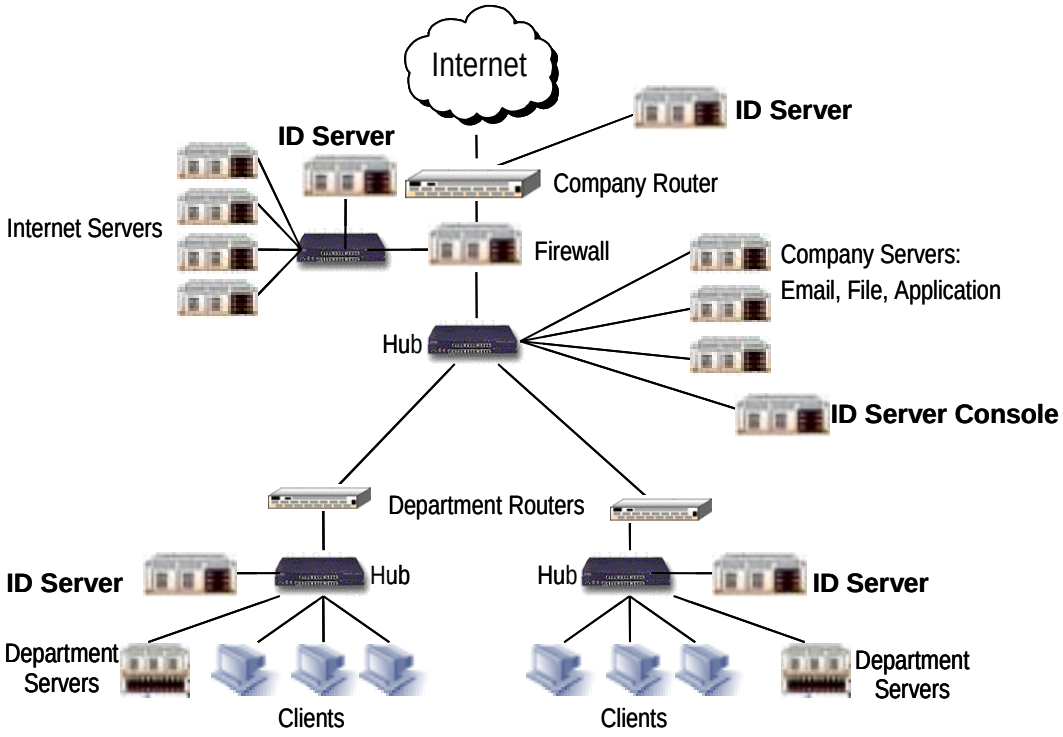
An intrusion detection system continually monitors network traffic, looking for known patterns of attack. When it detects an unauthorized activity, the system responds automatically with predetermined actions. It can report the attack, log the event, or terminate the unauthorized connection.

Intrusion detection software operates in concert with other security mechanisms to address the risks associated with hacking. Effective intrusion detection requires predicting vulnerabilities using tools that continually monitor the network for threats. Some obvious vulnerabilities can be the characteristic of a particular software or service.

Detecting an intrusion is equally as important as preventing one. Consider the locks on your office door and the alarm system that monitors the interior. The locks help prevent an intrusion, while the alarm system detects one. Ideally, no one ever breaks in; however, prudent security practices dictate that you should monitor for intrusions and alert the authorities upon the discovery of an intruder.

This same principle applies for your network security. Your security policy must consider the impact of an intrusion and provide a method for detecting it. When an intrusion occurs, the persons responsible for security in your organization must be alerted, so they can eliminate the vulnerability in your network security.

Intrusion Detection Server Placement



Intrusion detection servers can be placed in front of the firewall, behind the firewall, or run on the firewall itself.

One of the most effective places for an intrusion detection server is behind the firewall. An intrusion detection server behind the firewall covers most traffic destined for important servers, such as email servers, file servers, and web servers. Because it is located directly behind the Internet entry point, it also catches suspicious traffic if the firewall is compromised.

In a well-protected environment, intrusion detection servers should be implemented on each segment of the network. Intrusion detection servers implemented on internal network segments monitor employees' activities on the network.

ISS RealSecure Intrusion Detection

RealSecure is a real-time intrusion detection integrated solution. RealSecure monitors packet flow over a network in real-time and uses a rule-based approach to analyze packets for known attack patterns and unauthorized activity.

RealSecure improves security of an organization by:

- Providing an effective, second line of defense behind firewalls. This solution intercepts and disposes of attacks that get through the firewalls, and those that originate inside the firewalls.
- Providing a means of measuring the effectiveness of the current network security mechanisms. For example, RealSecure can detect when a Telnet session is being established from the Internet, even if the firewall has been configured to disallow Telnet sessions from the Internet.
- Providing feedback on vulnerabilities that have been accepted by the organization to provide access to a required service as identified by a vulnerability assessment tool. RealSecure can determine the number of attempts to exploit the vulnerability. If the number of attempted break-ins is unacceptable, the organization might choose to review the decision to make the service available.
- Providing useful reports to help organizations assess and tighten their security.

RealSecure Components

RealSecure is an automated, real-time intrusion detection and response system that unobtrusively analyzes activity across computer systems and networks. RealSecure includes two components:

- **Workgroup Manager (console)** — The user interface that controls the management of the sensors and reports network activity.
- **Sensors** — Monitor different types of suspicious activity that indicate attacks and respond appropriately when attacks are discovered. RealSecure uses two types of sensors: network sensors and server sensors. Using sensors together offers the best combination of early detection, rapid response, and detailed intruder tracking.

A daemon is a program that communicates alerts from the sensor to the console and from the console to the event collector. Daemon components include event collectors and sensors.

In RealSecure version 6.0 or later, managing and monitoring are separate actions. In previous versions of RealSecure, connecting to a sensor or daemon provided monitoring and managing capabilities.

Managing consists of connecting to a daemon or daemon component to modify its properties, modify its state, or apply policies to it. When an asset is being managed, it is listed in the Managed Assets window of the console. You must manage a daemon component before you can monitor it.

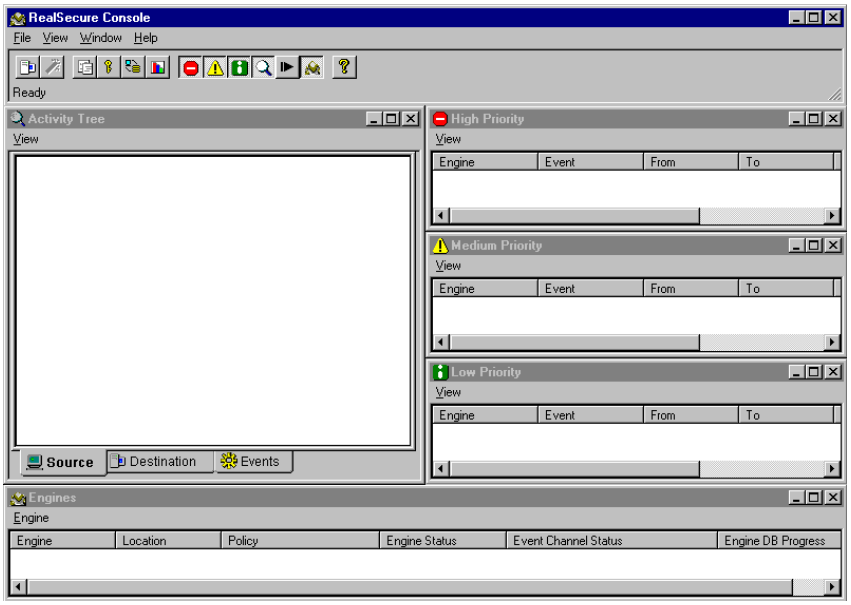
Monitoring consists of viewing detected attack events with the Display response enabled. When an asset is monitored, events related to that asset appear in the Priority windows. If you stop monitoring an asset, it is still active, but events related to the asset no longer appear in the Priority windows.

When RealSecure detects an attack, it reacts automatically according to its configuration. When an attack is detected, RealSecure can:

- Alert appropriate personnel through administrator console messages, email, or pager alerts.
- Log the event, along with associated information.
- Kill the event by terminating its connection.
- Initiate a user-supplied script.

Workgroup Manager

RealSecure provides components to aid in controlling deployed sensors, reporting attacks, and responding to attacks. The Workgroup Manager provides a user interface that consists of an asset database, event collector, and enterprise database. In addition, RealSecure provides a command line utility that is used to control sensor.



The Workgroup Manager provides a single point of management and control for all sensors in the network. An administrator can configure sensors, define policies, create signatures, and generate reports.

When a sensor detects an attack, it reports it to the Workgroup Manager. The Workgroup Manager displays the event in real-time optionally including the hacker's keystrokes and a copy of the hacker's screen. Attack events are classified and displayed by high, medium, or low priority.

Command Line Interface

The Command Line Interface (CLI) allows you to perform the following functions:

- Control RealSecure assets
- Receive alerts from RealSecure assets

The CLI is an alternative to the RealSecure console for managing RealSecure assets, and it operates in the same way as another session of the RealSecure console. You can use the CLI to run scripting commands and to schedule a control task such as a database upload.

Sensors

A RealSecure sensor monitors network and system traffic to detect attacks. Using sensors together provides the best combination of early detection, rapid response, and detailed intruder tracking.

Network Sensors

Network sensors monitor network packets and look for events that could indicate an attack against a network. Network sensors monitor all traffic on their network segments.

Server Sensors

Server sensors monitor network traffic for activities that indicate misuse or attacks. Log files and kernel-level activity are also monitored. Server sensors use intelligent alerting, block suspicious traffic, and intercept packets before they reach the operating system. Server sensors can monitor and control both inbound and outbound communications in this manner.

Sensor Deployment

For effective protection, deploy RealSecure network sensors on each network segment that contains a vital network or informational resource. RealSecure server sensors should be deployed on each critical server. For example, data resource servers containing critical customer information should have a server sensor to detect and terminate unauthorized access to this data.

RealSecure complements your existing authorization mechanisms by preventing and recording attempts to bypass these mechanisms.

Network Sensors

Common places to install RealSecure network sensors are as follows:

- In the *demilitarized zone* (DMZ)
- Just inside the firewall, on the intranet
- On key segments of the internal network

In the DMZ

By installing a network sensor in the DMZ of the Internet access point of a network, you can protect the devices installed in the DMZ from attack. Protecting the firewall is important, because the firewall acts as the control point for data flowing into your internal network and is often the initial target of an attack.

By adding a network sensor to your DMZ, you are dedicating an additional processor to the defense of your network perimeter. Every Internet access point should include a firewall and a network sensor.

Inside the Firewall on the Intranet

By installing a network sensor inside the firewall, you can detect changes to firewall operation and monitor the traffic passing through the firewall. A network sensor installed inside the firewall ensures that:

- The firewall is functioning properly, has not been compromised, and has not been misconfigured.
- The tunnels through the firewall are not being used to launch an attack against your internal network.

You can also use this network sensor in conjunction with a network sensor in the DMZ to evaluate the effectiveness of your firewall. For example, you could record the serious events at both network sensors and then generate a report of these events that compares the number of occurrences seen inside the firewall with the number seen outside.

Server Sensors

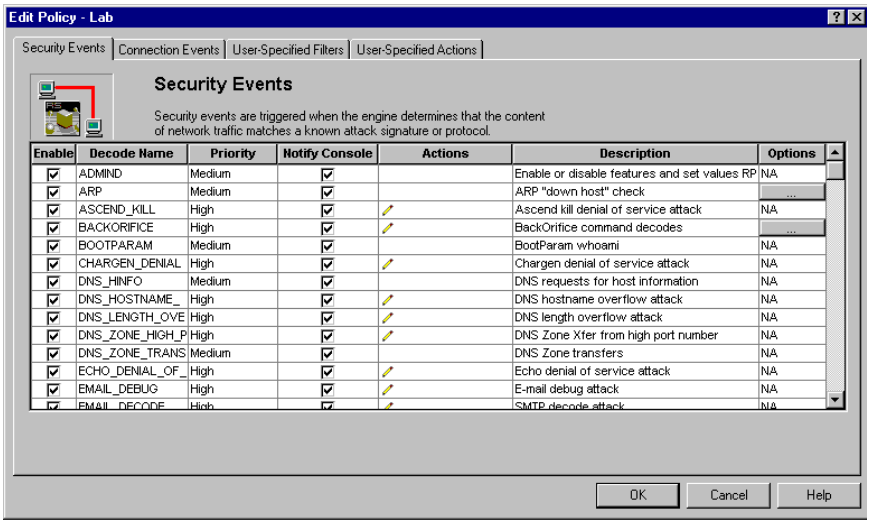
Common places to put RealSecure server sensors are:

- **Business-critical Windows servers** — By installing a server sensor on servers associated with vital applications or data files, the RealSecure sensors help monitor security-sensitive activities on the critical hosts.
- **Host systems with critical data** — By installing a server sensor on systems with sensitive data, you can detect changes in host configuration, unusual administrator activity, or attempts to access important files.
- **Windows domain servers** — Windows domain servers are typically the repository for important user account files and other important configuration data. The server sensor monitors activity on these critical data stores.

Deploying Multiple Sensors

RealSecure supports multiple sensors on the same computer. You can install one network sensor for each network segment as well as a server sensor on the same computer to provide additional protection and to concurrently monitor your network.

Configuration

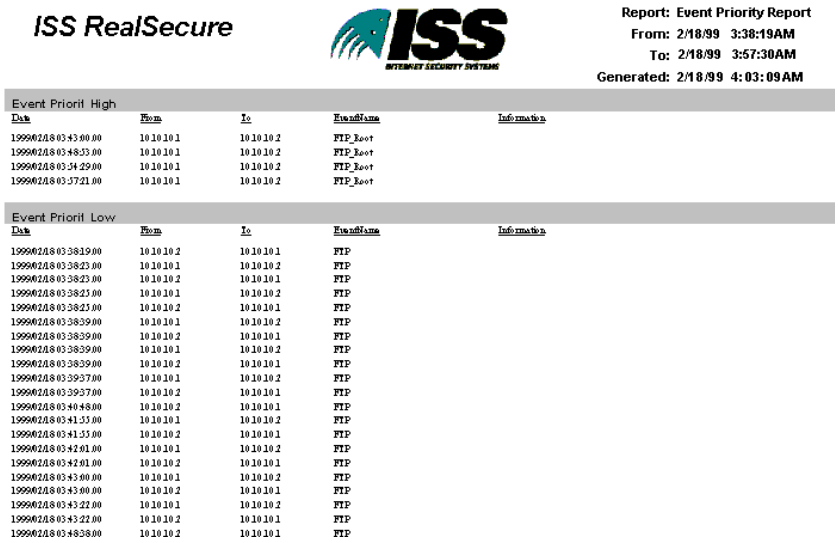


Sensors are configured through the Workgroup Manager interface. The configuration specifies the types of checks to be performed by the sensors and the response to be initiated for each type of attack detected.

Through the console, you can tailor a security model for the network to match the security policy of the organization.

RealSecure includes a default configuration to allow an organization to quickly implement security. The default configuration is biased toward tight security to ensure the protection of the monitored network. RealSecure also includes a variety of sample configurations, at different levels, to provide starting configurations for various types of security policies.

Reports



ISS RealSecure

Report: Event Priority Report
From: 2/18/99 3:38:19AM
To: 2/18/99 3:57:30AM
Generated: 2/18/99 4:03:09AM

Event Priorit High

Date	From	To	EventName	Information
19990218033430.00	1010101	1010102	FTP_Root	
19990218033833.00	1010101	1010102	FTP_Root	
19990218033429.00	1010101	1010102	FTP_Root	
19990218033721.00	1010101	1010102	FTP_Root	

Event Priorit Low

Date	From	To	EventName	Information
19990218033819.00	1010102	1010101	FTP	
19990218033823.00	1010101	1010102	FTP	
19990218033823.00	1010102	1010101	FTP	
19990218033825.00	1010101	1010102	FTP	
19990218033825.00	1010102	1010101	FTP	
19990218033839.00	1010101	1010102	FTP	
19990218033839.00	1010102	1010101	FTP	
19990218033839.00	1010101	1010102	FTP	
19990218033839.00	1010102	1010101	FTP	
19990218033937.00	1010101	1010102	FTP	
19990218033937.00	1010102	1010101	FTP	
19990218034048.00	1010102	1010101	FTP	
19990218034135.00	1010101	1010102	FTP	
19990218034135.00	1010102	1010101	FTP	
19990218034201.00	1010101	1010102	FTP	
19990218034201.00	1010102	1010101	FTP	
19990218034300.00	1010101	1010102	FTP	
19990218034300.00	1010102	1010101	FTP	
19990218034322.00	1010101	1010102	FTP	
19990218034322.00	1010102	1010101	FTP	
19990218034838.00	1010102	1010101	FTP	

RealSecure can generate meaningful reports from its event log files. These reports can include information such as the amount of data processed by a web server each day or the number of connections that were terminated each day and by whom. The Workgroup Manager can display these reports in graphical form, such as bar or pie charts, for easy review and analysis.

Administrators can use the reports to optimize security. The suggested method of implementation is to start with extremely tight filtering, using the default configuration. The administrator can use the information in the reports to tune filtering over time, as normal network activity is better understood. Repeated tuning reduces the number of alerts without relaxing security.

ISA Intrusion Detection

ISA Server includes intrusion detection, based on technology from ISS, which monitors for several well-known network vulnerabilities. ISA Server detects intrusions at two different network layers. First, ISA Server detects intrusions at the IP packet layer. This enables ISA Server to detect vulnerabilities that are inherent to the IP protocol. Second, ISA Server uses application filters to detect intrusions at the application layer.

Upon detecting an intrusion, ISA Server initiates a predefined alert. Alerts can be configured to send an email message, run a program, report to the event log, or start and stop services on the ISA Server. You can also define thresholds for alerts that prevent random detections that pose no immediate danger.

Intrusion Detection at the IP Packet Layer

The IP packet filtering engine in ISA Server provides intrusion detection for well-known IP vulnerabilities. An attacker can exploit IP vulnerabilities to scope information about a network, disable it, or trick a firewall into allowing access to the network.

The ISA Server IP packet filtering engine detects the following IP vulnerabilities:

- Windows out-of-band
- Land attack
- Ping of death
- Port scan
- IP half scan
- UDP bomb

Intrusion Detection at the Application Layer

ISA Server provides built-in support that detects DNS and POP protocol intrusions. You can create additional intrusion detection filters using the application filter interfaces in the ISA Server Software Development Kit.

The DNS intrusion detection filter detects the following known DNS vulnerabilities:

- **DNS hostname overflow** — An overflow occurs when the actual length of a message is greater than the expected length. The targeted computer is likely to fail or execute an undesired behavior. The DNS host name overflow occurs when a DNS server returns a hostname that is longer than the expected length.
- **DNS length overflow** — A DNS application expects a host name lookup response to contain a 4-byte length field. The DNS length overflow occurs when the length of the field is longer than 4-bytes.
- **DNS zone transfer from privileged ports (1 to 1024)** — A malicious user executes a zone transfer to gather a list of all the host names in a domain. In most cases, it is unwise to share the internal list of host names with Internet users. This filter detects when an Internet user attempts to execute a zone transfer using one of the privileged ports.
- **DNZ zone transfer from high ports (above 1024)** — This filter detects when an Internet user attempts to execute a zone transfer using one of the high (unprivileged) ports.

Reacting to Events with Alerts

After an event is triggered, ISA Server checks the list of alerts for an action to perform. ISA Server has predefined alerts for all event types. Alerts can be modified to perform the action appropriate for your network. The following options are configurable for each individual alert:

- **Send email** — You can send alert messages to a specific SMTP server, to primary recipients, and to carbon-copy recipients. You can also specify a reply address for the alert message.
- **Run program** — An alert can execute a program under the service account or under an account specified in the alert configuration.
- **Report to event log** — This option allows you to control whether an event log entry is created for the alert.
- **Start and stop services** — You can specify one or more of the ISA Server services to start or stop when an event occurs.

ISA Server allows you to specify when an alert is issued. You can specify the number of occurrences and the number of events per second before an alert is issued. After the initial alert, you can specify when the alert will be triggered a second time. The options include: immediately, after a manual reset, or after a specified amount of time in minutes.

Check Point Malicious Activity Detection

The FireWall-1 Malicious Activity Detection (MAD) feature provides a mechanism for detecting malicious or suspicious events and notifying the system administrator. The MAD feature runs on the management server and analyzes log files by matching log entries to attack profiles. The FireWall-1 administrator can modify attack detection parameters, turn detection on or off for specific attacks, or disable the MAD feature entirely.

The MAD feature scans the log file for the following attacks:

- Successive alerts
- Port scanning
- Blocked connection port scanning
- Login failure
- Successive multiple connections
- Land attack
- Local interface spoofing
- SYN attack
- Spoofed packets

Learning Check

1. In a network environment, where are the best places for implementing intrusion detection servers?

   .....

   .....

2. When RealSecure detects an attack, what are the possible responses?

   .....

   .....

   .....

3. What are the main components of ISS RealSecure?

   .....

   .....

4. What hardware subsystems impact performance of ISS RealSecure?

   .....

   .....

5. List the vulnerabilities detected by the ISA Server IP packet filtering engine.

   .....

   .....

   .....

   .....

   .....

   .....

   .....

6. How does the Firewall-1 MAD feature detect malicious or suspicious events?

   .....

   .....

   .....

Implementing Firewall Security

Module 9

Objectives

After completing this module, you should be able to:

- Describe a firewall, the terminology associated with firewalls, and the common features of firewalls.
- List and describe the key features of Microsoft ISA.
- Describe the architecture of ISA.
- List and describe the key features and architecture of Check Point VPN-1/Firewall-1.
- List the factors that contribute to firewall failures.

Firewall Overview

From enterprises that are web-enabling existing business applications and offering products and services through electronic storefronts, to corporations that are building secure extranets to streamline their supply chains, Internet applications are increasingly mission-critical. These applications, which companies are increasingly dependent on for revenues and customer-focused activities, have more stringent availability and scalability requirements than conventional systems.

Firewalls are the best practice for portal protection when private networks are connected to the Internet. Firewalls are most commonly placed between unknown or untrusted networks, such as the Internet, and known or trusted networks, such as a corporate network.

Typically, a firewall is hosted on a single computer that controls all traffic traveling between two networks and examines content as it passes through the firewall. This examination can be on inbound, outbound, or bidirectional traffic. Examination of the traffic content is based on a set of rules. Firewalls allow an administrator to create rules that specify the actions to be taken by the firewall on every packet it receives. These actions can include:

- Dropping the packet, if it is using an unauthorized protocol
- Accepting the packet and forwarding it to the appropriate destination
- Checking the packet for viruses
- Decrypting a packet, if a Virtual Private Network (VPN) is being used

Due to the nature of the function that firewalls perform, it is essential that the firewall maintain a detailed track of the state of each network connection. How firewalls keep track of state information and how much information they retain varies between those firewalls that implement stateful inspection and those that use application proxies.

Note

State information refers to the ability of a firewall to keep track of the status of all connections with and through the firewall. TCP connections have more state information associated with them than UDP connections.

Microsoft ISA Overview

ISA is an enterprise firewall and web cache server that integrates closely with Windows 2000 for policy-based security and straightforward management of internetworking. ISA Server provides three modes:

- A high-performance web cache server
- A multilayer firewall
- An integrated mode combining a firewall and cache

The cache improves network performance and end-user experience by storing frequently requested web information, while the multilayer firewall provides enterprise-class security. The firewall screens communication at the IP packet, circuit, and application layers. It controls access policy and the routing of network traffic. The cache and firewall can be deployed separately on dedicated servers or used together on the same server.

One of the key benefits of ISA Server is manageability. A difficult to manage firewall increases the potential for security vulnerabilities. ISA Server provides a rich set of easily managed security features through an MMC snap-in. The MMC snap-in provides easy to use security wizards and the ability to tweak security settings.

A common security approach is to restrict inbound connections from the Internet while allowing outbound connections for internal clients. This simplistic approach is problematic because a misused outbound connection is just as dangerous as an inbound connection.

Example

A nefarious program running on the internal network could connect to an external computer, allowing someone to execute commands remotely on the internal network.

ISA Server makes securing both inbound and outbound connections possible while allowing communication over the Internet. ISA Server provides this capability through the following features:

- Multilayer firewall
- Application filters
- System hardening
- Secure communication

Multilayer Firewall

A firewall that protects a network must provide three assurances:

- It must ensure that the information on the network is harmless. This includes protecting information on the network against unauthorized access. Additionally, the firewall must ensure that users on the network are accessing appropriate and safe information on the Internet.
- The firewall must ensure the privacy and accuracy of the information transferred to and from the network.
- The firewall needs to protect the reliability and availability of web and email servers and other network services.

ISA Server provides firewall features that fulfill each of these requirements by controlling access at several communication layers.

- IP packet filtering
- Circuit-level filtering
- Application-level filtering
- Dynamic packet filtering

Each layer has advantages and disadvantages that balance increased security against maximum performance. ISA Server makes the best use of all four firewall design strategies to provide an efficient and secure firewall.

IP Packet Filtering

IP packet filtering works by intercepting and evaluating packets before allowing passage through the firewall. Any packets allowed through must meet conditions specified by the administrator. If a packet is denied passage through the firewall, it is dropped. IP packet filters control access by source and destination IP addresses, communication protocol, and the IP port (communication channel) used.

Advantages	Disadvantages
Fast because little processing is required to inspect each packet and compare it to the rule set.	Packets are not inspected.
All packets are dropped unless explicitly permitted.	Packets are not manipulated they pass through the firewall.
No configuration is required on the client to process IP packet filters.	

Circuit-Level Filtering

When filtering at the circuit level, the firewall monitors communication sessions between computers on different networks. The firewall verifies a session by determining if a data packet is a connection request, belongs to a connection, or consists of a virtual circuit between two peer transport layers. If a session is valid, communication is permitted through the firewall.

A common example of this method is allowing internal clients to establish connections to the Internet using specific protocols such as FTP. The return connection completes the communication.

Advantages	Disadvantages
Faster than application filters because less processing needs to occur for each packet.	Packets are not inspected.
All packets are dropped unless explicitly permitted by a rule.	Caching is not available.
	Authentication is not available unless the request originates from a firewall client.

Application-Level Filtering

Access can also be controlled at the application level. By understanding the communication protocol used by an application, the firewall uses inherent intelligence to protect against dangerous or inappropriate communication. Only packets that comply with the definitions of the protocol are allowed passage through the firewall. Additionally, application-level firewalls can provide detailed session logs and user authentication.

Advantages	Disadvantages
Are intelligent enough to inspect the information passing through the firewall. Therefore, these filters can close application-specific vulnerabilities.	Some performance degradation is inherent due to the extra processing.
Firewall rules can specify application-specific resources to grant or deny access.	Each application type must have an application-level filter configured to inspect or proxy the communication.

Dynamic Packet Filtering

The final method of filtering data through a firewall is dynamic packet filtering. This type of filtering dynamically manipulates the firewall rules in response to user requests. When a client requests information through the firewall, IP packet filters are opened for the duration of the communication to allow a response from the server. When the communication ends, these temporary IP packet filters are removed. This is a highly secure method of filtering because only a minimum number of specific IP packet filters are open at any given time.

Advantages	Disadvantages
Have the same advantages as IP packet filters, but have the added advantage of allowing a minimum number of open IP filters at any given time	Packets are not inspected.
	Packets are not manipulated they pass through the firewall.

ISA Server uses a combination of all these filters in a hybrid solution that produces the best performance, provides the highest security, and includes the most features.

Application Filters

The ISA Server application filters offer the most intelligence in securing your network. Application filters intercept specific communication transferring through the firewall and inspect the actual payload or content of the packet. For example, the SMTP filter intercepts communication on port 25 and inspects it to ensure the SMTP commands are authorized before passing the communication to the destination server.

Application filters can serve three purposes.

- They can provide protection at the application layer by inspecting communication.
- They can redirect communication to another destination or create dynamic IP packet filters. For example, the HTTP Redirector filter redirects HTTP traffic to a Web Proxy Service.
- They can detect and alert the system of possible network intrusions.

ISA Server includes built-in application filters that are used to protect the network. ISA Server also includes an API for creating new application filters.

SMTP Filter

The SMTP Filter works by intercepting communication that arrives on port 25 and comparing the SMTP commands against a rule set. If the communication is allowed, it is passed to the destination server.

FTP Access Filter

The FTP access filter enables SecureNAT clients to use the FTP protocol secondary connections and allows you to restrict the use of the FTP protocol by adding protocol definitions to ISA Server. The FTP access filter creates protocol definitions that enable you to control what types of FTP communication are allowed through the firewall.

Enabling the FTP client protocol allows internal users to make FTP connections to Internet servers. You can use the FTP server protocol definition to create a server publishing rule that allows Internet users access to an internal FTP server.

Web Filters

ISA Server extends the security of the Web Proxy Service through web filters. You can create web filters by using the ISA Server Software Development Kit (SDK), or you can purchase them from third-party vendors. Web filters process all HTTP requests that pass through the Web Proxy Service. A web filter might inspect requests for viruses or filter inappropriate content.

System Hardening

ISA Server provides special system hardening wizards that make it easy to harden the operating system of the ISA Server. The system hardening wizards apply policy objects to the local server, reducing the number of services that are running and enabling security features inherent in Windows 2000. The following levels of security are available through the ISA Server security wizard:

- Use the secure setting when other applications are running on the computer (other than ISA Server). For example, you could use this setting if the IIS is also installed on the ISA Server.
- Use the Limited Services setting for servers running in integrated mode, operating both cache and firewall services.
- Use the Dedicated setting when only the ISA Server Firewall Service is running on the server. This is the most secure setting.

Secure Communication

ISA Server includes SSL tunneling and bridging. An SSL tunnel occurs when a web proxy client makes an HTTPS request for a resource on the Internet. After ensuring the connection is permitted, the Web Proxy Service opens a secure tunnel between the client and the server. The communication remains encrypted as it is transferred from the client, through the tunnel to the server.

SSL bridging occurs when ISA Server either encrypts or decrypts a request as it passes through the firewall. ISA Server can either end or initiate an SSL connection.

Example

SSL requests can be forwarded as HTTP requests in a reverse publishing scenario. In this case, the Internet user makes an SSL request to the ISA Server for an internally published website. Then, ISA Server decrypts the request and forwards it as an HTTP request to the internal server.

Internal clients can also benefit from SSL bridging. Typically, when a client makes an HTTPS request to a destination server, the ISA Server uses SSL tunneling, which establishes a direct connection between the server and client. For clients that support secure communication with the ISA Server, requests are first decrypted at the ISA Server. The ISA Server checks the cache to see if the resource is available locally. If it is not, the ISA Server makes a new request to the destination server. The new request may be HTTP or HTTPS depending on the routing rules.

ISA Architecture

ISA Server consists of several services, each providing different functions in the firewall architecture. These services are closely integrated with Windows 2000 providing a robust architecture. Depending on the options selected at installation, all or some of the services are present.

The ISA Server Control Service is a Windows 2000 service that controls starting and restarting of the other ISA Server services. It manages synchronization of configuration information in ISA Server Arrays, updating client configuration files and deleting unused log files. The Control Service generates alerts and executes actions in response to server security and health conditions.

The ISA Server Firewall Service responds to requests from Firewall and SecureNAT clients. It acts as a secure router, routing permitted IP traffic between the external and internal networks. The Firewall Service also uses an application filter to intercept HTTP requests and redirect them to the Web Proxy service. This service is present only when ISA Server is installed in firewall or integrated mode.

The Web Proxy Service handles requests from HTTP, FTP, and Gopher clients. It makes requests to origin servers on behalf of clients. The Web Proxy Service caches content locally and determines when the requests are served from the local cache. This service is present when ISA Server is installed in cache or integrated mode.

The Scheduled Cache Content Download Service downloads HTTP content into the cache on a schedule predetermined by the administrator. When in the cache, requests for the content can be retrieved from the local network instead of transporting them across the upstream link. This service is present when ISA Server is installed in either cache or integrated mode.

Rules for ISA Server Security

Rules instruct ISA Server to accept and process requests from internal and external web clients in a specified manner.

- **Packet filter** — Packet filters control the types of IP packets accepted on the external interface. Packet filters are available in Integrated and Firewall mode only. When enabled, all packets on the external interface are dropped unless a specific packet type is configured to be accepted.

  Typically, you create packet filters to control incoming traffic or traffic that is routed between the perimeter network and external network. For outgoing requests, ISA Server opens ports dynamically as they are needed and monitors the ports for responses.

- **Web publishing rules** — Web publishing rules are used to configure ISA Server to forward requests from external clients to internal web servers. These rules are used in a reverse caching scenario to accelerate access to web servers within an organization. Web publishing rules determine security restrictions for incoming requests, how the requests are encrypted, and when they are forwarded to the internal server.

- **Bandwidth rules** — ISA Server uses bandwidth rules to determine what priority to request from the QoS service. You can use these rules to specify a priority to real-time communications or to a group of users who depend on Internet access.

- **Protocol rules** — Protocol rules control access to specific protocols that are allowed to pass through the ISA Server. In cache mode, the only protocols available are HTTP, HTTPS, FTP, and Gopher. In integrated and firewall mode, any protocols can be defined using protocol definitions and allowed or denied using protocol rules.

- **Site and content rules** — Site and content rules control access to specific destination servers and content types by internal clients. Access is controlled using destination sets, which can be an entire domain, a specific server, or a specific URL. Content restrictions are first defined using content groups and then restricted using site and content rules.

- **Routing rules** — Routing rules route requests to upstream ISA Server computers or redirect requests to alternate destination servers. Routing rules specify which requests are routed, redirected, or retrieved directly from the destination server.

  Routing rules are used to configure caching policies specific to destination sets. Additionally, routing rules specify whether to serve objects from the cache and whether to cache the responses from the destination server.

Rules Processing Order

You can group ISA Server rules into two types of rule sets, ordered and unordered. ISA Server processes the rules differently depending on the type of rule set.

Ordered rule sets include web publishing rules, bandwidth rules, and routing rules. ISA Server processes these rule sets in the order they are listed. The last rule in an ordered rule set is the default rule and cannot be deleted. When a request arrives related to one of these rule sets, ISA Server will use the first rule in the list that matches the request.

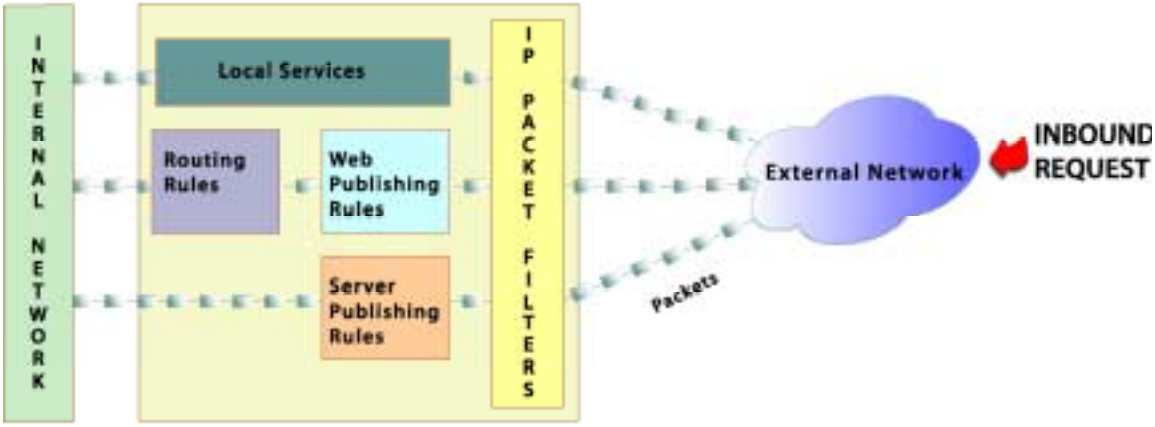
Unordered rule sets include packet filters, protocol rules, and site and content rules. For these rule sets, ISA Server processes "deny" rules first and then processes "accept" rules.

The default for unordered rule sets is to deny everything. This means that if no rules exist in a rule set, ISA Server will deny all requests for that rule set.

Incoming Requests

Incoming requests are processed in the following order:

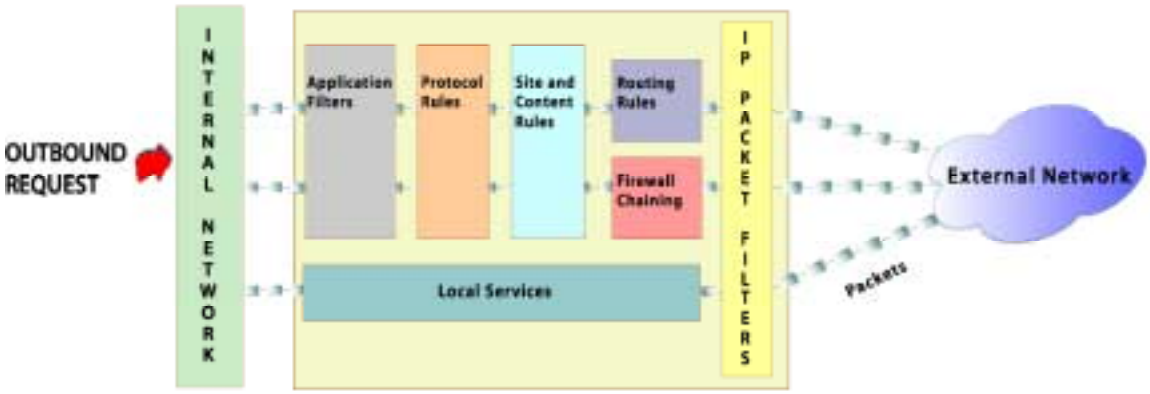
1. Packet filters
2. Web publishing rules and server publishing rules
3. Routing rules



Outgoing Requests

Outgoing requests are processed in the following order:

1. Application filters
2. Protocol rules
3. Site and content rules
4. Routing rules and Firewall Chaining
5. Packet filters



Check Point VPN-1/Firewall-1 Overview

Check Point VPN-1/FireWall-1 is a comprehensive security platform that provides:

- Access control
- User authentication
- Network address translation (NAT)
- Virtual private networking (VPN)
- Content security (anti-virus, URL and Java/ActiveX screening)
- Auditing and reporting
- Lightweight Directory Access Protocol (LDAP)-based user management
- Intrusion detection
- Malicious activity detection
- Third-party device management
- High availability and load sharing

VPN-1/Firewall-1 provides organizations with the ability to define an integrated security policy that can be distributed across multiple firewall gateways and managed remotely.

Check Point stateful inspection technology examines every packet passing through key locations in your network (Internet gateway, servers, hosts, routers or switches), promptly blocking all unwanted communication attempts. Packets do not enter the network unless they comply with the enterprise security policy. An auditing mechanism centralizes logs and alerts from the entire system at the management console.

VPN-1/FireWall-1 is completely transparent to both users and applications, and coexists with other security tools.

VPN-1/Firewall-1 Architecture

The VPN-1/FireWall-1 architecture enables an organization to define and implement a centrally managed security policy. The enterprise security policy is defined at a central management console and downloaded to multiple enforcement points throughout the network.

VPN-1/FireWall-1 consists of the following components:

- Check Point Policy Editor Graphical User Interface (GUI)
- Management Server
- VPN/FireWall Module

Check Point Policy Editor

The Check Point Policy Editor GUI enables you to define policies in terms of network objects (hosts, networks, and gateways) and rules. A VPN-1/FireWall-1 security policy is defined in terms of a rule base and properties.

Six kinds of policies can be defined:

- Security Policy — Specifies the types of communications allowed to enter and leave the network and how connections will be authenticated and encrypted.
- Network Address Translation Policy — Specifies how invalid internal IP addresses will be translated to valid IP addresses.
- Quality of Service (QoS) Policy — Specifies the allocation of bandwidth resources among connections, maximizing throughput.
- Desktop Security Policy — Enables you to control access to desktops remotely and within the local network.
- WebAccess Policy — Enables you to manage authorization requirements for web applications.
- VPN Manager Policy — Enables you to manage VPN communities.

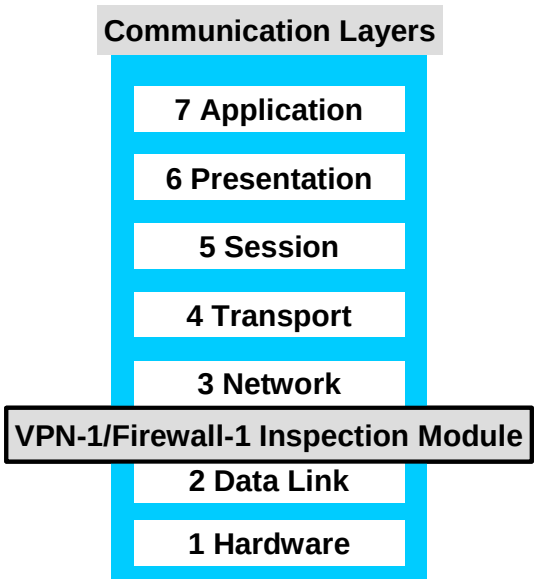
Each object defined in the Policy Editor has a set of attributes, such as name or IP address. Only network objects that are explicitly referenced in a policy are defined.

A rule base is an ordered set of rules against which each communication is checked. Each rule specifies the source, destination, service, and action to be taken for each communication such as whether it is permitted or denied. A rule also determines how communication is tracked such as by sending an alert message.

Management Server

Policies defined using the Policy Editor GUI are saved on the Management Server. The Management Server maintains the Check Point databases, including network object definitions, user definitions, policies, and log files for enforcement points.

VPN-1/FireWall-1 Inspection Module



The VPN-1/FireWall-1 Inspection Module resides in the operating system kernel, between the Data Link and the Network layers (layers 2 and 3) of the OSI model. Because the data link is the actual network interface card and the network link is the first layer of the protocol stack, VPN-1/FireWall-1 is positioned at the lowest software layer.

Inspecting at this layer ensures that the VPN-1/FireWall-1 Inspection Module intercepts and inspects all inbound and outbound packets on the gateway. The higher protocol stack layers do not process packets unless the Inspection Module verifies that they comply with the security policy. VPN-1/FireWall-1 examines IP addresses, port numbers, and any other information required to determine whether packets should be accepted in accordance with the security policy.

VPN-1/FireWall-1 accesses and analyzes data derived from all communication layers. Any traffic not explicitly allowed by the security rules is dropped by default and real-time security alerts are generated.

Firewall Failure

Firewalls have traditionally been implemented without any means of redundancy, with the possible exception of a standby server that must be switched over manually to assume the role of the primary server. This manual switchover can lead to several hours of downtime. Therefore, this manner of providing portal security does not work well with Internet business applications, since downtime translates into lost revenue.

Firewall vendors have taken different approaches to providing state information that can be shared among firewalls for the purpose of failover. Some vendors provide as much state information as possible. Other vendors intentionally provide less state information. These vendors view failover events as security risks. To minimize this risk, their approach is to revalidate all connections to ensure that the initial reason for the firewall failure was not due to a malicious attack, which could be carried over to the backup firewall if all state information was retained.

When selecting a solution, each company must evaluate these risks against the customer service aspects of a complete and seamless failover. The decision depends on the unique situation of each company, protocols used, connectivity, systems behind the firewall, and the existing firewalls. No single solution will satisfy the needs of all companies.

Firewalls may fail due to several factors, including:

- Hardware failures of the processor, disk, memory, and power supply.
- Operator error
- Failure of the operating system due to lack of system resources or errors in coding
- Failure of the firewall software due to lack of resources or errors in coding
- Denial of service attacks by hackers or disgruntled employees

Failover may not be successful if the cause of the failover is related to software or to specific attacks. If the primary firewall has problems in these areas and fails, there is a strong likelihood that the secondary firewall will inherit the same issues and may also fail. Because these issues require manual interpretation, an effective management and notification system for firewall failures is important.

Maintenance Events

While most people think of high availability in terms of a failure event, it is more likely that the primary reason for a firewall not being operational is actually a planned maintenance event. A firewall must be disabled periodically to perform system maintenance, upgrade the operating system software, upgrade the firewall software, and to change allowed protocols and scanning features. Therefore, it is essential to have the capability of performing an orderly transfer to a backup firewall, and then to restore the initial firewall to the operational state after the upgrades have been made.

Learning Check

1. What is the purpose of a firewall?

   .....

   .....

2. List the features used by Microsoft ISA to secure both inbound and outbound connections.

   .....

   .....

3. List the ISA Server services and briefly describe their functions in the firewall architecture.

   .....

   .....

   .....

4. List the components of Check Point VPN-1/Firewall-1.

   .....

   .....

   .....

5. Explain why an effective management and notification system for firewall failures is important.

   .....

   .....

   .....

   .....

   .....

   .....

Implementing Virus Protection

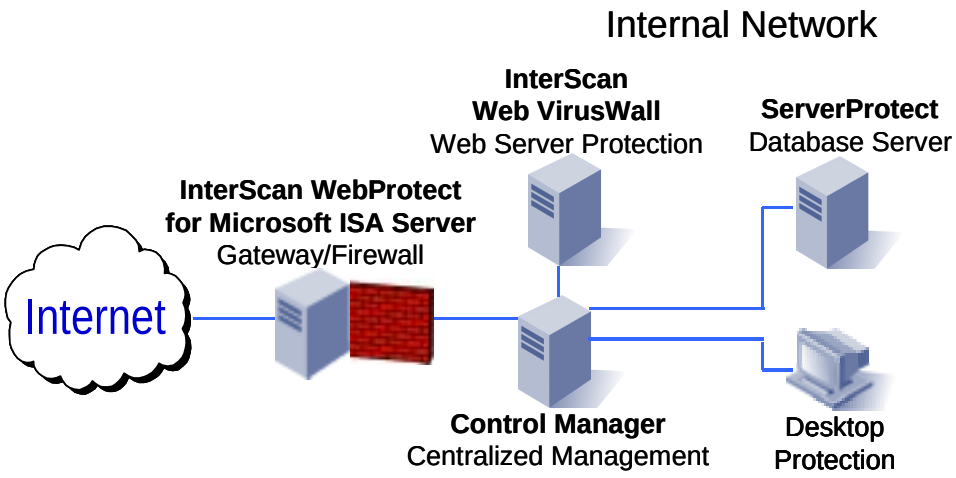
Module 10

Objectives

After completing this module, you should be able to:

- Describe the function and key features of Trend Micro ServerProtect.
- Describe the function and key features of Trend Micro InterScan WebProtect for Microsoft ISA Server.
- Describe the function and key features of Trend Micro InterScan VirusWall.
- Describe the function and key features of Trend Micro Control Manager.

Trend Micro Product Overview



HP and Trend Micro have partnered together to ensure optimal integration and performance with these Trend Micro products:

- ServerProtect
- InterScan WebProtect for Microsoft ISA Server
- Control Manager
- InterScan VirusWall

ServerProtect Features

Trend Micro ServerProtect provides server-based virus protection that can detect known and unknown viruses, including even unknown macro viruses. It is designed to protect the network server with minimal performance and administration overhead. ServerProtect detects new file infections, identifies viruses in existing files, and detects activity indicating the presence of an unknown virus in the network environment—either on the server, workstation, or both.

ServerProtect has the following features:

- **Centralized domain management** — Provides antivirus protection of multiple Windows 2000 servers and domains from a single management console. Servers in the same domain can be configured simultaneously from a single console and virus incident reports can be generated from all servers.

  With a means to configure, monitor and maintain antivirus efforts through a management console, ServerProtect improves and simplifies the implementation of corporate virus policy.

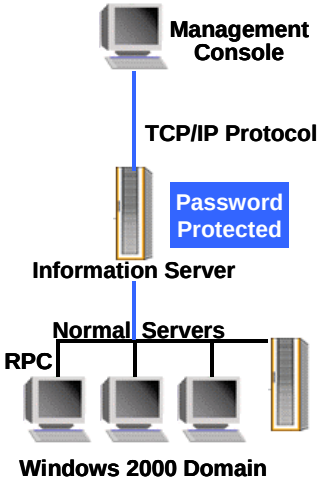
- **Central Quarantine Tool** — The Central Quarantine Tool quickly and safely moves files quarantined by ServerProtect to any normal server on the network. For example, if there are 10 normal servers with quarantined files, use the Central Quarantine Tool to move all the quarantined files to a single, central normal server folder.

- **ScanNow** — The ScanNow tool performs a manual scan on a normal server. Use this tool to immediately scan the normal server without having to use the management console.

- **Trojan System Cleaner** — The Trojan System Cleaner detects the activity of live Trojan horse programs, kills Trojan processes, deletes the files dropped by the processes, and recovers system files that have been modified by Trojans.

- **Task-oriented operation** — Tasks can be created and scheduled for operations such as Scan Now, update, print logs, export logs, purge logs, run statistics, and real-time scanning.
- **Scheduled virus pattern updates** — ServerProtect Information Server can be configured to download virus pattern and scan engine updates on a regular basis and then distribute them to designated normal servers.
- **Rollback** — ServerProtect enables the rollback of virus pattern, scan engine, and program files after performing an update. If ServerProtect suffers stability problems or false alarms after a component update, roll back the update and restore the software to its original state.
- **Flexible control over infected files** — When ServerProtect detects that a file is infected, the file can be restored after cleaning, sent to Trend Micro if it is still suspect or uncleanable, and returned to the user through email after cleaning.
- **Automated response to virus outbreaks** — ServerProtect can provide notification using a message box, pager, SMTP message, SNMP trap, printer, and Windows NT event log to warn of virus infection, service load or unload, real-time scanning configuration changes, deny write intrusion, and virus pattern out-of-date.
- **Log management** — ServerProtect provides a comprehensive log report that records antivirus activities on the network. The events include virus infection, pattern or program updates, virus alerts, running tasks, scanning activities, modifications to the write-protected files, and file update results.

ServerProtect Architecture



ServerProtect enables secure management of the entire network through a three-tier architecture. The management console uses TCP/IP protocol with password-protected logon to the ServerProtect Information Server. The information server administers other ServerProtect clients, called normal servers, using Remote Procedure Call (RPC) to connect to Windows 2000 servers.

Normal Server

A normal server can be any server in a network on which ServerProtect is installed. In the ServerProtect architecture, the normal server is the first line of defense against viruses and is managed by an information server.

Information Server

The information server is a communications hub for coordinating antivirus defense activities within its domains. It provides a single point of contact for the normal servers assigned to it, simplifying the task of directly communicating with each normal server.

The number of servers an information server can manage is only limited by the available bandwidth because it is simply a delivery system for information. Limiting the number of normal servers assigned to an information server simplifies management.

Management Console

A ServerProtect Management Console is a portable console that runs on a Windows platform to centrally control multiple network servers and domains.

Note

Only one management console can manage an information server. After an information server has been connected to a management console, no other consoles can connect to that information server.

How ServerProtect Works

ServerProtect monitors all activity on Windows 2000 networks. Whenever it detects that a file in its domain is being accessed, it checks the file for infection.

If it finds that the file has a virus, it sends notification messages to predefined recipients and acts on the virus according to its configuration. ServerProtect records all system activity.

How ServerProtect Finds Viruses

Using a process called pattern matching, ServerProtect searches an extensive database of virus patterns to identify known virus signatures. Key areas of suspect files are examined for distinctive strings of virus code and compared against thousands of virus signatures that Trend Micro has on record.

For polymorphic, or mutation viruses, the ServerProtect scanning engine permits suspicious files to execute in a protected area within which it is decrypted. ServerProtect then scans the entire file, including the newly decrypted code, and looks for strings of mutation-virus code. If a virus is found, ServerProtect performs a virus action such as clean, delete, bypass (ignore), remove, or rename.

Macro viruses are not associated with a particular operating system because they are application specific; the application acts as their operating system. Given this cross-platform compatibility, the growing popularity of the Internet and email, and the increasing power of macro languages, the magnitude of the threat posed by these viruses is obvious. The Trend Micro MacroTrap is designed to provide a means of protecting the network users from receiving and sending macro viruses.

The MacroTrap performs a rule-based examination of all macro code that is saved in association with a document. Macro virus code is typically contained as a part of the invisible template that travels with the document. MacroTrap checks the template for signs of a macro virus by seeking out instructions that perform virus-like activity. Examples of virus-like activity are copying parts of the template to other templates (replication) and copying code to execute harmful commands (destruction).

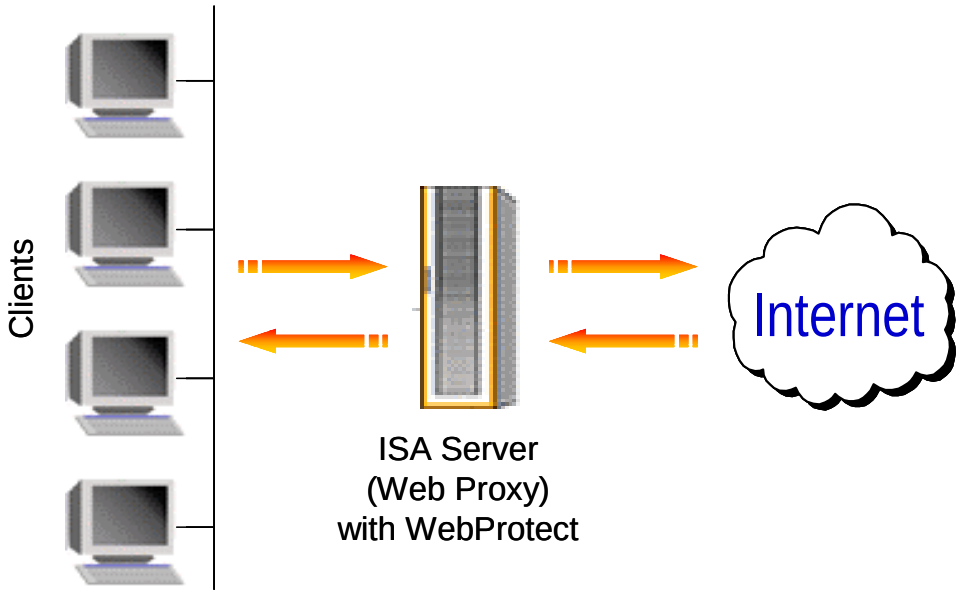
Trend Micro InterScan WebProtect for Microsoft ISA Server

InterScan WebProtect is a set of programs that function as a gateway between a LAN and the Internet. InterScan WebProtect scans all or selected file transfer through HTTP for viruses. The program performs a user-configured action such as clean, quarantine, delete whenever it encounters an infected file.

Following are some of the features of InterScan WebProtect:

- Integrates with firewall applications.
- Intercepts viruses at the server before they enter and damage your network.
- Uses a Windows 2000-based virus scanning engine.
- Allows you to filter MIME content that would otherwise be obstructed by the scanning function.
- Provides an easy web browser configuration.
- Transparent to the user.
- Provides periodic, prescheduled virus pattern file updates.

How InterScan WebProtect Works



InterScan WebProtect examines all incoming files transferred through HTTP, places the contents into a temporary file, and invokes the virus-checking program. If the incoming file is not infected, InterScan WebProtect passes the file to the client.

If a file is infected with a virus, InterScan WebProtect performs a predefined action. It may delete the file, pass it to the recipient node, or isolate it for later action.

InterScan WebProtect uses a virus pattern file to detect known signature viruses by pattern matching. In addition to catching known signature viruses, the program detects and intercepts previously unknown polymorphic or mutation viruses.

When the InterScan scanning engine detects polymorphic or mutation viruses, it saves the virus code in a temporary location on the WebProtect server. After decrypting the code, InterScan WebProtect characterizes the mutation virus and performs a preconfigured action such as clean or quarantine.

InterScan WebProtect uses the Trend Micro generic macro virus scanning engine, MacroTrap, to detect macro viruses. MacroTrap is a rule-base virus detection method, which detects and removes known as well as unknown macro viruses.

InterScan VirusWall Overview

InterScan VirusWall is a gateway solution that monitors inbound and outbound email, web transactions, and FTP traffic for viruses. InterScan is largely "set and forget," which means that after it is installed and configured, all routine tasks can be scheduled to occur automatically. A variety of plug-ins, called eManager, are also available, which provide control over the allocation of network bandwidth, enterprise-wide junk mail filtering, and content filtering.

Depending on the speed of the machine and the volume of network traffic, each VirusWall can be installed on a different server or all three VirusWalls can be installed on the same server.

For large network environments such as a WAN, where multiple instances of InterScan will be installed, or networks running several different Trend antivirus products, administration can be consolidated and managed collectively using Control Manager.

How InterScan VirusWall Works

InterScan employs an extensive database of virus signatures, inert snippets of virus code, and a set of heuristic rules to quickly compare and analyze the binary and macro content of files for viruses. This database is commonly called the virus pattern file.

INTERNET Trend Micro publishes new virus pattern files weekly. These updates are available at **http://www.trendmicro.com**.

In most cases, it is recommended to install InterScan on a server that is inside the firewall and in front of the server or proxy server that it will complement. In other words, after a firewall, InterScan should be first in line to receive the network SMPT, HTTP, or FTP traffic. It can then monitors all inbound SMTP, HTTP, and FTP for viruses.

When it detects a file type that it has been configured to scan (for example, .zip, .exe, .doc), InterScan places the file into a temporary location and opens it for virus checking. Key areas of the files are checked for matching virus signatures against the Trend Micro database of thousands of known viruses. In addition, heuristic logic is used to analyze whether certain unknown code (unique to polymorphic viruses, for example) is virus-like in nature and warrants testing.

If the file being scanned is virus-free, it is routed to the original SMTP, FTP, or HTTP server, or firewall, which then delivers the file in the normal manner. InterScan can include a notification with the file indicating that it was scanned and found to be virus-free, or it can operate invisibly in the background.

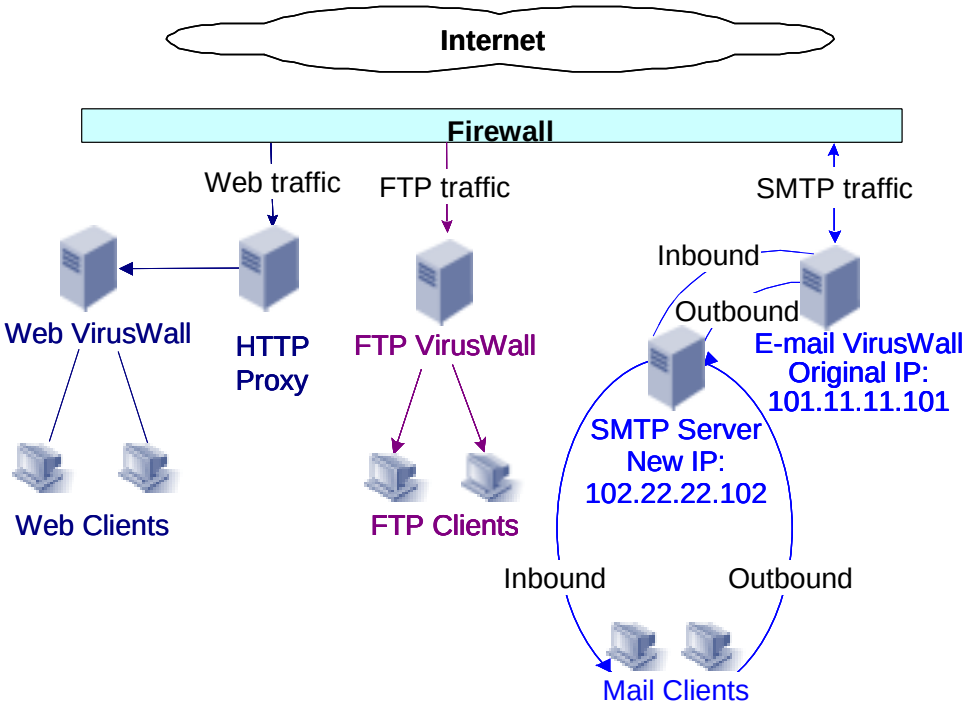
Virus Pattern Updates

Typically, administrators choose silent virus pattern file updates, but the operation can also be performed manually, remotely, or even outside of InterScan.

Silent updates occur automatically, for example each Tuesday at 4:00 a.m., and no user intervention is required. Manual updates can be performed either locally or remotely by the administrator by clicking the *Update Now* button.

Virus pattern files can also be downloaded to any machine from the Trend website and then placed, by hand, in the InterScan directory. In all cases, no special installation is necessary.

Implementing InterScan VirusWall

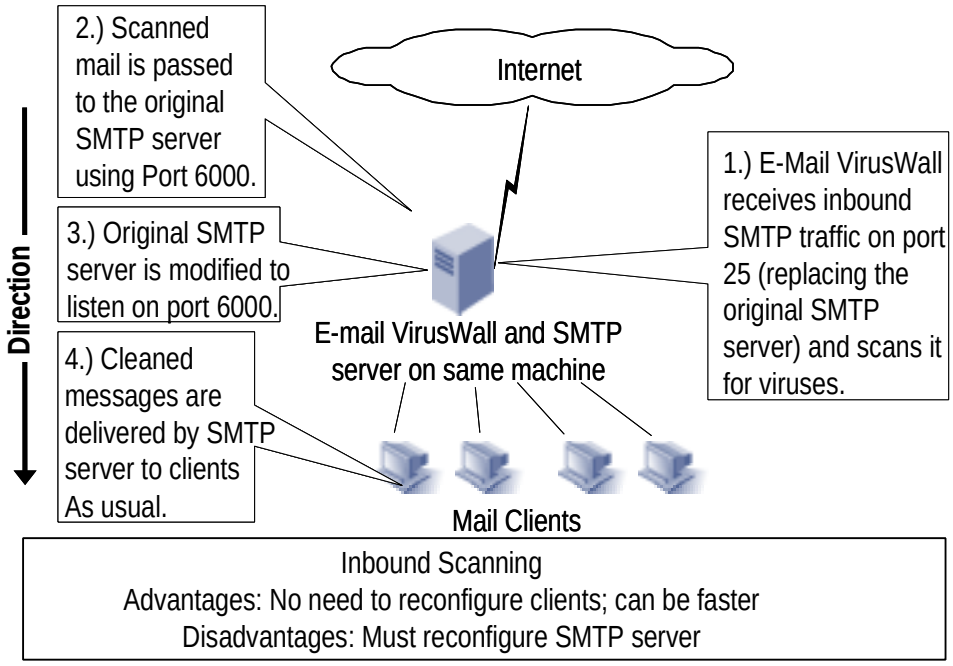


Depending on where InterScan is installed, you might need to modify your firewall configuration (IP or port) so that it routes traffic to InterScan. You might also need to modify the configuration of your SMTP, FTP, or web servers.

InterScan receives inbound and outbound Internet traffic as it crosses the gateway, checks for viruses, cleans infected files, and then sends the traffic to the original server, the requesting client, or the remote destination for delivery as usual.

By default, InterScan will use port 21 for FTP VirusWall, port 8080 for Web VirusWall, and port 25 for its E-mail VirusWall. In some cases, such as when there is an existing FTP server on the machine where InterScan is being installed, for port conflicts may occur. Port conflicts must be resolved at the time of initial configuration because no port can be shared between two different programs.

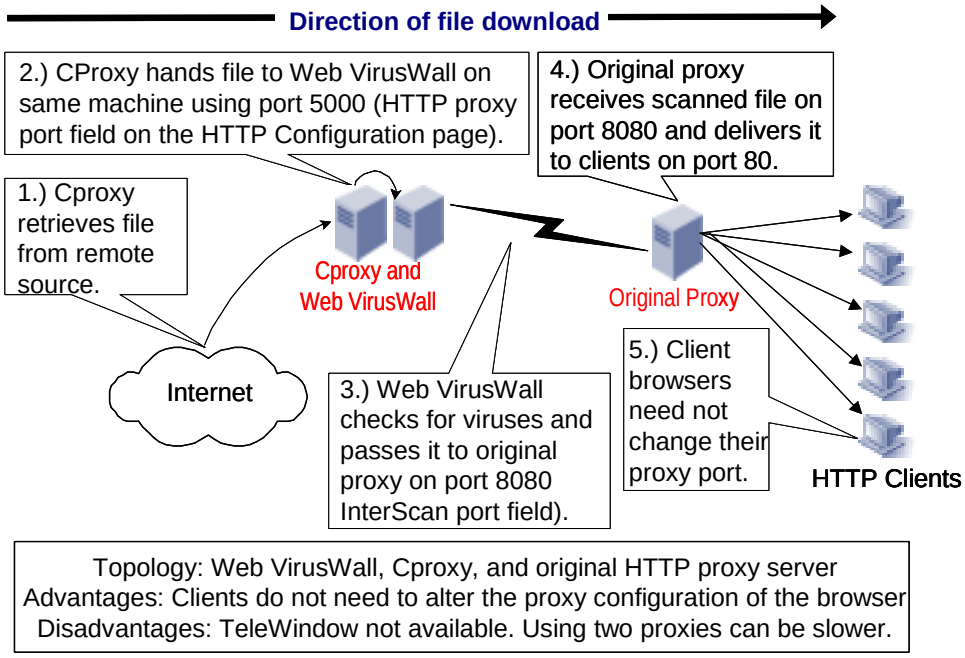
E-mail VirusWall



The InterScan E-mail VirusWall is designed to support most SMTP servers and can be installed on the same machine as the SMTP server or a different one.

For example, if you are running Microsoft Exchange server on the same machine where InterScan is being installed, you must change the port Exchange uses from 25 to something else. InterScan automatically detects Exchange and will prompt you to switch during installation.

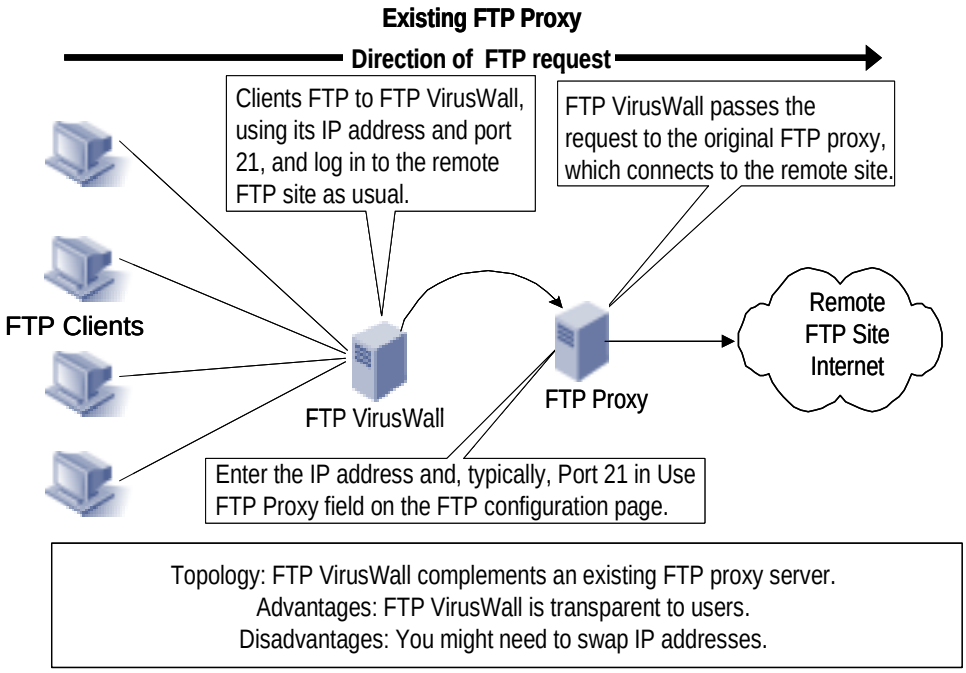
Web VirusWall



InterScan Web VirusWall requires a proxy server and is designed to support most HTTP proxy servers. If no proxy server exists on the network, one is optionally available with Web VirusWall. In addition to virus scanning, Web VirusWall allows you to apply a single Authenticode standard to all network users and block potentially harmful Java applets from being downloaded to the LAN.

Web VirusWall can be installed on the same machine as the proxy or a different one. In setting up Web VirusWall to work with an existing proxy server, the primary consideration is that the two proxy servers must not be configured to use the same port address. One setup topology is shown in the preceding graphic.

FTP VirusWall



FTP VirusWall is usually installed so that it will scan FTP file transfers that are being downloaded to the local LAN (as opposed to scanning such transfers for remote users who are accessing an FTP site that you are hosting). If you are already running an FTP proxy server on port 21, you need to stop the service before installing FTP VirusWall or change the port it is using. One such example is shown in the preceding graphic.

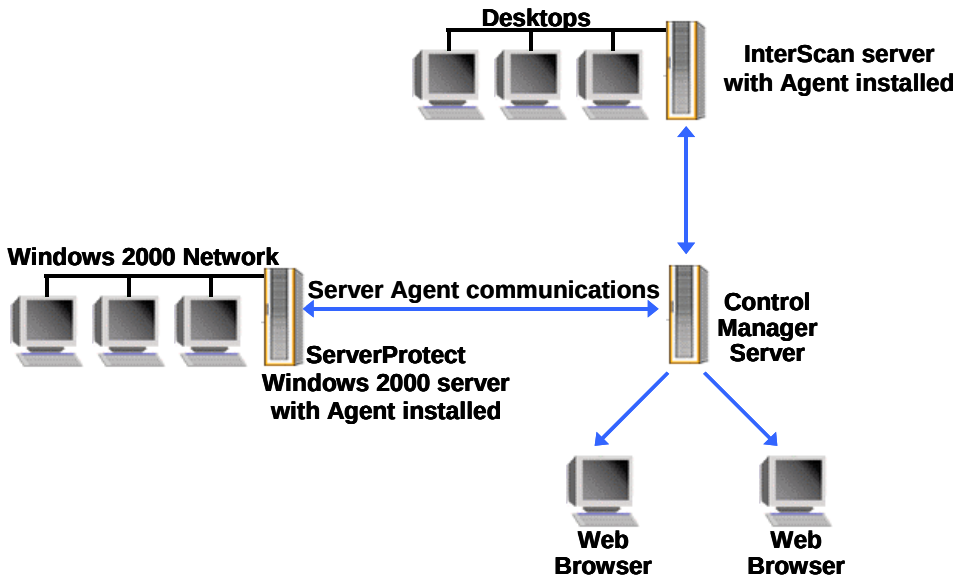
Trend Micro Control Manager

Trend Micro Control Manager is a management solution that enables you to configure, monitor, and maintain all antivirus programs installed on the network from a single point, regardless of the actual physical location or platform of the program. As part of the Trend Micro Enterprise Solution concept, Control Manager improves and simplifies the administration of corporate virus control policy.

Control Manager is a web-based tool that provides the following features:

- **Outbreak Commander** — Performs a wizard-like function that eliminates the guesswork in configuring your antivirus and content security defenses during a virus outbreak.
- **Proprietary communication security infrastructure** — Provides a communications infrastructure that is built on the SSL protocol. Depending on the security settings used, transmissions can be encrypted, or encrypted with authentication.
- **Task delegation** — Allows each user to be assigned a personalized account with its own privileges, which defines what the user can see and do. Account usage can be tracked via user logs.
- **Command tracking** — Allows you to monitor all commands executed on Control Manager. This is particularly useful for determining whether or not commands like virus pattern updates for example were performed successfully.
- **Support for antivirus and content security products** — Allows management of both antivirus and content security products.
- **Real-time product control** — Allows you to make configuration changes, which are immediately applied to the products.
- **Centralized installation of agents** — Permits agents, which are the primary interface between Control Manager and its managed products, to be installed directly on the target machine or remotely from the Control Manager console.
- **Centralized update control** — Permits centralized updates of your virus pattern and scan engine files to ensure that all products contain the latest components. It also enables you to view the protection status your entire network from a single management console.
- **Centralized configuration** — Allows you to coordinate virus-response and security efforts from a single console to ensure consistent enforcement of virus and security policies.
- **Centralized log reporting** — Allows you to review the performance of your antivirus network using comprehensive logs. Control Manager can collect logs from all its managed products, which eliminates the need to check the logs of each product.

Control Manager Architecture



Control Manager has a three-tier system:

- Client machine
- Product-level management server
- Control Manager server

The workstations, desktop machines, and servers are client machines and form the first tier. All virus protection occurs in the first tier.

The second tier is a product-level management layer. The Trend Micro client-server products, InterScan and ServerProtect, have management servers that control the different client machines.

A single control server in the third tier governs client-server products. Using the Control Manager web-based management console, you can enforce a uniform antivirus policy for all your antivirus products and platforms on the network.

Control Manager controls the first two tiers using applications called agents. The agents receive command inputs from the management console and apply them to the managed products. The management console obtains status information from the products and sends them back to the Control Manager server, allowing you to quickly view the status of your entire network.

Control Manager Components

Control Manager includes several components that are used to centrally manage antivirus strategies deployed on a network. The components are accessed within the web-based tool.

Control Manager Server

Control Manager Server is a Windows 2000 server running Windows 2000 and IIS. HTML pages, Java applets, and CGI code that generate the user interface are stored on the server.

The server also houses two databases. The virus event database stores centralized virus event information. The other database is used only if LDAP is not available. It stores configuration and domain information as well as information from the machines running antivirus software.

Directory Manager

Directory Manager allows you to design a directory tree to reflect the organizational infrastructure of a company. The infrastructure is usually based on the security policy of your company and defines how antivirus and content security policies are deployed and managed.

The Directory is the means by which managed products (entities) are controlled. It is a virtual organization of the different antivirus servers managed by the Control Manager server. Use the Directory Manager to organize the Directory in a way that supports your management topology.

User Manager

User Manager allows you to create account types and assign access rights. The access rights provide the ability to monitor assigned parts of the network, execute commands such as `scan now`, configure antivirus and security content policies, and modify the organization of the Directory Manager folders and entities.

Update Manager

The Update Manager facilitates obtaining the latest virus pattern and scan engine to ensure that your system can counter the latest threats. When creating a deployment plan, you must create the desired schedules first. A schedule is needed for each targeted group that is represented by a folder in the Directory Manager.

You can customize deployment times to determine which systems get updated first. This ensures that the gateway and email servers are updated before desktop environments. In addition, you can prioritize policy deployment to help reduce bandwidth use and provide faster deployment.

Outbreak Commander

Outbreak Commander addresses the entire outbreak life cycle. Outbreak Commander consists of three components:

- **Outbreak Prevention Services (OPS)** — Delivers proactive policy recommendations prior to the availability of pattern files. OPS helps to identify and contain outbreaks.
- **Pattern Files and Scan Engines** — Help to eliminate the virus completely from the network.
- **Damage Cleanup Services (DCS)** — Delivers an automated method of removing dangerous virus remnants from clients. DCS helps simplify the costly manual process of cleaning systems one by one.

Before to an attack, use the Outbreak Commander wizard within Control Manager to determine how pattern files are to be distributed to maximize efficient use of corporate bandwidth. You can set an appropriate virus pattern, scan the engine download schedule, and prepare an attack management deployment plan for efficient utilization of network bandwidth. You can also use Control Manager to configure Trend Micro antivirus products available on the network and to set manual scanning to optimize the response to an outbreak.

Control Manager Communication

Control Manager works in real time and communicates with Trend Micro products installed on your servers through agents. Agents communicate with the Trend server through a secure HTTP connection and then translate the HTTP commands into API calls that direct the local antivirus software.

The Agent Install program first detects which servers have antivirus software programs installed, then copies the agent software corresponding to the detected program to the server where it resides.

Example

Agent Install detects the antivirus program, InterScan VirusWall for 2000, on the server “Washington.” It automatically installs the agent that corresponds with InterScan on that server and registers itself with Control Manager. This data appears in the server tree on the Control Manager console.

After the agents are installed on servers running antivirus programs, the relationship between the agents and Control Manager is typically client/server. The agent takes on the dual role of gathering program-specific data from the local machine and handling the communications with the remote Control Manager server.

Learning Check

1. Explain how ServerProtect fits in an enterprise virus protection policy.

   .....

   .....

   .....

   .....

2. How does InterScan WebProtect detect known signature viruses?

   .....

   .....

   .....

3. Explain the purpose and benefit of InterScan VirusWall.

   .....

   .....

   .....

   .....

   .....

   .....

4. Explain the purpose and benefit of Control Manager.

   .....

   .....

   .....

   .....

   .....

Module 1 Learning Check

1. Describe the basic principles of the DISA framework.

   The DISA framework is largely based on conventional scalability and availability principles and technology, but applies these technologies in new ways to fit the Internet application paradigm.

2. List the components of the DISA.

   The Compaq DISA framework consists of a three core components and four global components. The core application stack components are access and acceleration resources, web and application servers, and data resources. The global components are security, operations and management, integration and services, and business applications.

3. What are the steps for deploying a DISA implementation on a modular basis?

   1. Implement a single server.
   2. Move the content to a dedicated server.
   3. Add load balancing.
   4. Add web and application servers.
   5. Add redundancy and capacity.
   6. Eliminate single points of failure.

4. What Microsoft components provide services at the web server layer of the DISA framework?

   Microsoft Content Management Server and Microsoft Commerce Server

Module 2 Learning Check

1. You are planning an intranet website for an advertising firm. The remote locations will access the intranet over a slow WAN connection. The web content will change often include many images. What are the considerations in planning the interface?

   Avoid the use of unnecessary graphics, which can increase it takes to download a web page.

2. Discuss the implications of manually building web pages for an intranet.

   If you choose to build a traditional “page-at-a-time” website, there is less planning required and it facilitates faster deployment. However, the drawback to this method is frequency of updates as well as maintenance costs.

Module 3 Learning Check

1. List the components within the Commerce Server architecture.

   The Business Analytic System, Business Processing Pipeline Systems, Business Desk, Commerce Server Manager, and the Commerce Server databases.

2. What is the purpose of a pipeline?

   A pipeline is a software framework that defines and links stages of a business process, running them in sequence to complete a specific task. Commerce Server includes the following pipelines that perform the processing required by an online store.

3. You want to collect information about users that visit your website. Which Commerce Server feature provides that functionality?

   Profiling System

4. After you develop your website, you need to package and move your site from the development environment to the test environment. Which Commerce tool provides this functionality?

   Commerce Server Site Packager

5. Which Commerce Server report would allow you to analyze the content users are seeking on your site?

   Query String Reports

6. Which process isolation option provided the best performance for the DISA configuration in HP testing?

   Medium

Module 4 Learning Check

1. List the six main functional components of Microsoft Content Management Server.

   Site Builder, Web Author, Server Configuration Application, Site Deployment Manager, Site Stager, and Content Connector.

2. A human resources staff member, who is assigned the role of moderator, needs to make a change to a job posting but is unable to access the page content. What is the problem?

   Even though a moderator can revise the schedule of a posting before approving it, they cannot revise the page content.

3. You need to add several new white papers to your website. Which Content Management Server component would you use?

   Site Builder

4. When a website is ready to be deployed into the content production environment, what application would you use to move the entire site?

   When a website is ready to be deployed into the content production environment, use SQL database transfer (or backup and restore) to move the entire site. For any subsequent incremental changes, use the Site Deployment Manager to move site structure, templates, and content.

Module 5 Learning Check

1. What are the four main factors that affect web server performance?

   Processor, Memory, Network, and Disk.

   The type of content served by the website also affects performance.

2. What are the three main capacity planning metrics?

   Requests per second, throughput, and processor utilization

3. What are the HP recommendations for capacity planning in relation to sizing memory?

   Have enough memory in the system to reduce or eliminate paging.

4. It seems that a lengthy processor queue forms when servicing a large number of connections. What can you do ensure that accepted connections are processed promptly?

   Limit connections so the server blocks or rejects connections during heavy traffic.

5. What System Monitor counter can be used to monitor the page file?

   Memory → Available bytes

6. List three memory optimization procedures that you think will improve the performance of your web server.

   Keep related web files on the same logical partitions of a disk.

   Have enough physical memory to hold all static web pages.

   Lengthen the period an unused object can remain in the cache.

Module 6 Learning Check

1. What are the three types of commercial IP load balancing products?

   Software-based, Ethernet-based, application-based

2. Into which of these three categories does NLB fit?

   Application-based

3. What process is initiated when the state of the server array changes?

   NLB servers periodically exchange multicast or broadcast messages within the array

4. What method does NLB use to associate requests from specific clients to specific servers?

   Convergence

5. How do NLB servers coordinate their actions?

   Affinity

Module 7 Learning Check

1. List three software vulnerabilities that account for the majority of attacks.

   - Default installs of operating systems and applications
   - Accounts with no passwords or weak passwords
   - Nonexistent or incomplete backups

2. What is the HP recommendation for securing SNMP?

   HP recommends not exposing unprotected computers running SNMP outside of the private network and changing the default community strings.

3. List the important security layers and briefly describe their functions.

   Physical security — The physical location of the server

   Network security — The logical locations of the server and the application-level protocols running over the network

   Host security — The security of the host server operating system

   Application security — The security of the web server application

   Website data security — The security of the website data

4. What are the disadvantages of a web server outside of the firewall configuration?

   The web server is not protected by the firewall against spoofing and protocol filtering; content inspection; IP activity logging; and network address translation. This exposes the web server to many types of network and protocol attacks. Also remote management and monitoring from inside the firewall is difficult and potentially insecure.

5. What is host security?

   Host security refers to the security features built into the server operating system.

6. What type of user authentication is required for a secure web server?

   The web server administrator should require secure user authentication using HTTPS over a security protocol such as SSL.

Module 8 Learning Check

1. In a network environment, where are the best places for implementing intrusion detection servers?

   One of the best places for an intrusion detection server is behind the firewall. This configuration covers most traffic destined for critical servers and catches suspicious traffic if the firewall is compromised.

2. What are the main components of ISS RealSecure?

   Workgroup Manager and sensors

3. When RealSecure detects an attack, what are the possible responses?

   Alert appropriate personnel through console messages, email, or pager alerts; log the event; kill the event; or initiate a user-supplied script.

4. If you deploy multiple sensors, what is the maximum number of network and server sensors that can be installed on a computer?

   You can install one network sensor for each network segment as well as a server sensor on the same computer.

5. List the vulnerabilities detected by the ISA Server IP packet filtering engine.

   ISA Server IP packet filtering detects Windows out-of-band, land attack, ping of death, port scan, IP half scan, and UDP bomb.

6. How does the Firewall-1 MAD feature detect malicious or suspicious events?

   The MAD feature analyzes log files by matching log entries to attack profiles.

Module 9 Learning Check

1. What is the purpose of a firewall?

   Typically, a firewall controls traffic traveling between two networks and examines content on inbound, outbound, or bidirectional traffic based on a set of rules.

2. List the features used by Microsoft ISA to secure both inbound and outbound connections.

   ISA Server secures both inbound and outbound connections using a multilayer firewall, application filters, system hardening, and secure communication.

3. List the ISA Server services and briefly describe their functions in the firewall architecture.

   ISA Server Control Service is a Windows 2000 service that controls starting and restarting of the other ISA Server services.

   ISA Server Firewall Service routes permitted IP traffic between the external and internal networks.

   Web Proxy Service handles requests from HTTP, FTP, and Gopher clients.

Scheduled Cache Content Download Service downloads HTTP content into the cache on a predetermined schedule.

4. List the components of Check Point VPN-1/Firewall-1.

   VPN-1/FireWall-1 consists of the Check Point Policy Editor Graphical User Interface (GUI), Management Server, and VPN/FireWall Module.

5. Explain why an effective management and notification system for firewall failures is important.

   If the primary firewall has software problems, there is a strong likelihood that the secondary firewall will inherit the same issues and may also fail. Because these issues require manual interpretation, an effective management and notification system for firewall failures is important.

Module 10 Learning Check

1. Explain how ServerProtect fits in an enterprise virus protection policy.

   With a means to configure, monitor and maintain antivirus efforts through a management console, ServerProtect improves and simplifies the implementation of corporate virus policy.

2. How does InterScan WebProtect detect known signature viruses?

   InterScan WebProtect uses a virus pattern file to detect known signature viruses by pattern matching.

3. Explain the purpose and benefit of InterScan VirusWall.

   InterScan VirusWall is a gateway solution that monitors inbound and outbound email, web transactions, and FTP traffic for viruses. InterScan is largely “set and forget,” which means that after it is installed and configured, all routine tasks can be scheduled to occur automatically.

4. Explain the purpose and benefit of Control Manager.

   Trend Micro Control Manager is a management solution that enables you to configure, monitor, and maintain all antivirus programs installed on the network from a single point, regardless of the actual physical location or platform of the program.

Glossary

Buffer overrun — A condition that results from adding more information to a buffer than it was designed to hold. An attacker may exploit this vulnerability to take over a system.

Daemon — A program that acts as an intermediary between the RealSecure console and sensor or event collector.

Demilitarized Zone (DMZ) — A network set up separately from the private network of an organization and the Internet. The advantage of a DMZ is that it allows external users access to specific servers located in the perimeter network, while preventing access to the internal corporate network.

Hacker — A person who works to penetrate the security barriers of a communications network. There are generally two types: malicious hackers, network intruders engaged in the illicit theft of corporate or customer information such as credit cards, and malevolent hackers, those who exploit network vulnerabilities as part of their work for security assessment firms.

Internet Server Application Programming Interface (ISAPI) — An interface designed by Microsoft to allow Windows applications to interact with Microsoft's Web server, Internet Information Services (IIS).

Microsoft Data Engine (MSDE) — A client/server data engine that provides local data storage on a smaller computer system, such as a single-user computer or small workgroup server. Unlike SQL Server, MSDE has a 2-gigabyte database size limit, does not support Symmetrical Multiprocessing, and cannot be a replication publisher in a replicated-database environment.

Netscape Server Application Programming Interface (NSAPI) — An interface designed by Netscape to allow applications to interact with Netscape Web servers.

NTLM — The NTLM protocol authenticates users and computers based on a challenge/response mechanism. When the NTLM protocol is used, a resource server must contact a domain authentication service on the domain controller for the computer or user's account domain to verify its identity whenever a new access token is needed.

LAN Manager (LM) is the least secure form of challenge/response authentication, but it is more secure than cleartext. NTLM version 1 is a more secure form of challenge/response authentication than LM. NTLM version 2 is the most secure form of challenge/response authentication.

Open Systems Interconnect (OSI) model —

Proxy server — A firewall component that manages Internet traffic to and from a LAN and can provide other functions, such as document caching and access control.

Round Robin DNS — A method of configuring a Domain Name Service (DNS) server so that DNS lookups of a particular host name are sequentially distributed across a pool of IP addresses instead of a single address. Each successive client visiting a site is directed to a different application server, thus achieving a rudimentary load-balancing effect.

Although RRDNS is inexpensive and straightforward to implement, there is no mechanism that enables the DNS server to receive feedback about the current load and availability of the application servers. If a server is overloaded or offline due to a failure, the DNS server will unwittingly continue to send clients to the failed server until an administrator manually updates the DNS configuration.

In response to the deficiencies of RRDNS, many vendors have developed intelligent load-balancing products. These products automatically detect failure of a server in the array and instantaneously route traffic to another server. Load-balancing technology currently comes in several forms, including software that is installed on a general purpose server such as Windows 2000 Network Load Balancing (NLB), turnkey networking appliances (such as Cisco LocalDirector), and switches.

Secure Sockets Layer (SSL) — A protocol for establishing an encrypted communications channel to help prevent the interception of critical information, such as credit card numbers on the World Wide Web and other Internet services.

Sniffer — A tool used to capture information going over a network.

Spoofing — The practice of making a transmission appear to come from a user other than the user who performed the action.

Trojan horse — A computer program that appears to be useful but that actually does damage.

Virtual private network (VPN) — A private data network that makes use of a public network, such as the Internet, by encrypting data at one node and using security procedures that provide a "tunnel" through which the data can pass to another node.

Virus — A program that attaches to other programs, which executes when those programs run. When a person runs an infected program, the virus infects other files on that computer and any system in which that computer is connected. If a computer is connected to a server, the server can become infected.

Worm — A program that propagates by attacking other machines and copying itself to them. The difference between a worm and a virus is the amount of human interaction involved and how it spreads from machine to machine. A worm does not have to be run to spread from machine to machine.