Deploying Cisco Wide Area Application Services

Second Edition

Design and deploy Cisco WAN optimization and application acceleration solutions for the enterprise WAN

Zach Seils, CCIE® No. 7861
Joel Christner, CCIE No. 15311
Nancy Jin

ciscopress.com

Deploying Cisco Wide Area Application Services, Second Edition

Joel Christner, Zach Seils, Nancy Jin

Published by:
Cisco Press
800 East 96th Street
Indianapolis, IN 46240 USA

Printed in the United States of America

First Printing January 2010

Library of Congress Cataloging-in-Publication data is on file.

ISBN-13: 978-1-58705-912-4

ISBN-10: 1-58705-912-6

Warning and Disclaimer

This book is designed to provide information about deploying Cisco Wide Area Application Services (WAAS). Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied.

The information is provided on an "as is" basis. The authors, Cisco Press, and Cisco Systems, Inc. shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or from the use of the discs or programs that may accompany it.

The opinions expressed in this book belong to the author and are not necessarily those of Cisco Systems, Inc.

Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Cisco Press or Cisco Systems, Inc., cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Corporate and Government Sales

The publisher offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales, which may include electronic versions and/or custom covers and content particular to your business, training goals, marketing focus, and branding interests. For more information, please contact: **U.S. Corporate and Government Sales** 1-800-382-3419 corpsales@pearsontechgroup.com

For sales outside the United States please contact: **International Sales** international@pearsoned.com

Feedback Information

At Cisco Press, our goal is to create in-depth technical books of the highest quality and value. Each book is crafted with care and precision, undergoing rigorous development that involves the unique expertise of members from the professional technical community.

Readers' feedback is a natural continuation of this process. If you have any comments regarding how we could improve the quality of this book, or otherwise alter it to better suit your needs, you can contact us through email at feedback@ciscopress.com. Please make sure to include the book title and ISBN in your message.

We greatly appreciate your assistance.

Publisher: Paul Boger

Associate Publisher: Dave Dusthimer

Executive Editor: Mary Beth Ray

Managing Editor: Patrick Kanouse

Senior Development Editor: Christopher Cleveland

Project Editor: Ginny Bess Munroe

Editorial Assistant: Vanessa Evans

Cover Designer: Sandra Schroeder

Book Designer: Louisa Adair

Composition: Mark Shirar

Cisco Representative: Erik Ullanderson

Cisco Press Program Manager: Anand Sundaram

Copy Editor/Proofreader: Deadline Driven Publishing

Technical Editors: Jim French, Jeevan Sharma

Indexer: Angie Bess

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Foreword

I am pleased to write the foreword to the second edition of *Deploying Cisco Wide Area Application Services (WAAS)*. Over the past few years, WAN Optimization technology has become a standard component of enterprise networks. The benefits accruing from the use of the technology for server consolidation, simplified IT management, and improvement of the efficiency of information sharing and network utilization have earned it a place at the top of customers' buying priorities.

At Cisco, we have made several innovations to our award-winning WAAS solution that continues to expand the benefits it offers our customers. These include the use of virtualization technology—that is, Virtual Blades (VB)—to rapidly deploy a network service "anytime, anywhere," and a variety of application specific acceleration techniques that we developed in collaboration with the leading application vendors.

At Cisco, we believe that WAN optimization technology needs to be closely integrated with the routing/VPN architecture of the enterprise network so that customers can benefit from a single, optimized, shared network fabric that delivers all applications: voice, video, and data.

The authors combine experience from their work with thousands of customers who have deployed large installations of WAAS with a deep knowledge of enterprise and service provider network design, IOS, application-aware networking technologies, and WAAS to provide a comprehensive set of best practices for customer success. I strongly recommend customers who are interested in WAN optimization and particularly Cisco WAAS to read this volume. It will help you accelerate your understanding of the solution and the benefits you can accrue.

George Kurian
Vice President and General Manager, Application Networking and Switching
Cisco Systems, Inc.

Introduction

IT organizations are realizing the benefits of infrastructure consolidation and virtualization—cost savings, operational savings, better posture toward disaster recovery—and the challenges associated. Consolidating infrastructure increases the distance between the remote office worker and the tools they need to ensure productivity—applications, servers, content, and more. Application acceleration and WAN optimization solutions such as Cisco Wide Area Application Services (WAAS) bridge the divide between consolidation and performance to enable a high-performance consolidated infrastructure.

This book is the second edition of *Deploying Cisco Wide Area Application Services*, and updates the content to reflect the innovations that have been introduced in version 4.1.3 of the Cisco Wide Area Application Services (WAAS) solution, whereas the first edition was written to version 4.0.13. Along with coverage of the key components of the Cisco WAAS solution, this edition expands on the concepts introduced in the first edition to provide a more complete understanding of the solution's capabilities, how to use them effectively, and how to manage them. This edition expands upon the first edition to include coverage for new solution components including application-specific acceleration techniques, hardware form factors, virtualization, application performance management (APM), monitoring and reporting enhancements, and workflow enhancements. Additional technical reference material is provided in the appendices to help familiarize users of version 4.0 with changes that have occurred in the command-line interface (CLI) with the introduction of the 4.1 release. A quickstart guide is provided to help users quickly deploy in a lab or production pilot environment in order to quantify the benefits of the solution. A troubleshooting guide can also be found at the end which helps associate difficulties encountered with potential steps for problem resolution.

Goals and Methods

The goal of this book is to familiarize you with the concepts and fundamentals of sizing and deploying Cisco WAAS in your environment. The book provides a technical introduction to the product, followed by deployment sizing guidelines, through integration techniques, and configuration of major components and subsystems. The intent of the book is to provide you with the knowledge that you need to ensure a successful deployment of Cisco WAAS in your environment, including configuration tips, pointers, and notes that will guide you through the process.

Who Should Read This Book?

This book is written for anyone who is responsible for the design and deployment of Cisco WAAS in their network environment. The text assumes the reader has a basic knowledge of data networking, specifically TCP/IP and basic routing and switching technologies.

As the WAAS technology continues to evolve, the content in this book will provide a solid framework to build on. Mastering the topics in this book will ensure that you can approach any WAAS design project with confidence.

How This Book Is Organized

Although this book could be read cover to cover, it is designed to be flexible and allow you to easily move between chapters and sections of chapters to cover just the material that you need to work with . Although each of the chapters builds upon the foundation laid by previous chapters, enough background information is provided in each chapter to allow it to be a standalone reference work in and of itself. Chapter 1 provides a technical examination of the Cisco WAAS product and its core capabilities, along with use cases and the "why you care" about each of the solution components. Chapters 2 through 10 are the core chapters and, although they can be covered in any order, it is recommended that they be covered sequentially for continuity. Chapter 11 provides a series of use cases for the Cisco WAAS product family, which can also provide insight into how other customers use this technology to meet their business infrastructure requirements. Appendices are provided to help augment and also summarize what is discussed in the core chapters. Following is a description of each chapter:

- **Chapter 1, "Introduction to Cisco Wide Area Application Services (WAAS):"** This chapter provides a technical examination and overview of Cisco WAAS and its core components.
- **Chapter 2, "Cisco WAAS Architecture, Hardware, and Sizing:"** This chapter discusses the Cisco WAAS appliance and router-integrated network module hardware family, positioning of each of the platforms, and system specifications that impact the design of a solution relative to the performance and scalability of each component.
- **Chapter 3, "Planning, Discovery, and Analysis:"** Planning is a critical part to any successful WAAS deployment. Spending ample time at the beginning of the project to understand the requirements, including those imposed by the existing network environment, is critical for a successful deployment. Chapter 3 gives you a head start by outlining the key topic areas that should be taken into consideration as you are planning your WAAS deployment.
- **Chapter 4, "Network Integration and Interception:"** This chapter provides an in-depth review of the network integration and interception capabilities of Cisco WAAS. The topics discussed in Chapter 4 form the foundation for the design discussions in subsequent chapters.
- **Chapter 5, "Branch Office Network Integration:"** This chapter provides a detailed discussion of the different design options for deploying Cisco WAAS in the branch office environment. Several design options are discussed, including detailed configuration examples.
- **Chapter 6, "Data Center Network Integration:"** This chapter examines the key design considerations for deploying WAAS in the data center. Sample design models and configuration examples are provided throughout the chapter. Best practices recommendations for scaling to support hundreds or thousands of remote sites are also included.

- **Chapter 7, "System and Device Management:"** This chapter walks you through the initial deployment of the Central Manager and each of the accelerator WAAS devices, including the setup script, registration, federated management, and use of management techniques such as device groups. This chapter also provides a detailed understanding of integration with centralized authentication and authorization, alarm management, an introduction to the monitoring and reporting facilities of the CM, CM database maintenance (including backup and recovery), and the XML-API.
- **Chapter 8, "Configuring WAN Optimization:"** This chapter guides you through the WAN optimization framework provided by Cisco WAAS, including each of the optimization techniques and the Application Traffic Policy manager. This chapter also examines the configuration of optimization policies, verification that policies are applied correctly, and an examination of statistics and reports.
- **Chapter 9, "Configuring Application Acceleration:"** This chapter focuses on the application acceleration components of Cisco WAAS, including configuration, verification, and how the components interact. This chapter also looks closely at how these components leverage the underlying WAN optimization framework, how they are managed, and an examination of statistics and reports.
- **Chapter 10, "Branch Office Virtualization:"** This chapter examines the virtualization capabilities provided by certain Cisco WAAS appliance devices, including configuration, management, and monitoring.
- **Chapter 11, "Case Studies:"** This chapter brings together various topics discussed in the previous chapters through several case studies. The case studies presented focus on real-world deployment examples, a discussion of the key design considerations, options, and final device-level configurations.
- **Appendix A, "WAAS Quickstart Guide:"** Appendix A provides a quickstart guide to help you quickly deploy WAAS in a proof-of-concept lab or production pilot.
- **Appendix B, "Troubleshooting Guide:"** Appendix B provides a troubleshooting guide, which helps you isolate and correct commonly encountered issues.
- **Appendix C, "4.0/4.1 CLI Mapping:"** Appendix C provides a CLI mapping quick reference to help identify CLI commands that have changed between the 4.0 and 4.1 versions.

Chapter 2

Cisco WAAS Architecture, Hardware, and Sizing

Chapter 1, "Introduction to Cisco Wide Area Application Services (WAAS)," introduced the performance challenges created by the wide-area network (WAN) and how they are addressed by the Cisco WAAS solution. Cisco WAAS is a software component that is resident on a hardware device deployed at each location with users and servers. This hardware device, which can be deployed as a router-integrated network module for the Integrated Services Router (ISR) or as an appliance, is named either Cisco Wide-Area Application Engine (WAE) or Cisco Wide-Area Virtualization Engine (WAVE). The distinction between the two is that a WAVE device, available only as an appliance, can also provide branch office virtualization services in conjunction with WAN optimization and application acceleration. WAE devices provide only WAN optimization and application acceleration and do not provide virtualization.

This chapter provides an introduction to the Cisco WAAS hardware family, along with an in-depth examination of the hardware and software architecture. This chapter also looks at the licensing options for Cisco WAAS, positioning for each of the hardware platforms, and performance and scalability metrics for each of the platforms.

Cisco WAAS Product Architecture

The Cisco WAAS product family consists of a series of appliances and router-integrated network modules that are based on an Intel x86 hardware architecture. The product family scales from 512 MB of memory to 24 GB of memory, utilizing single-processor subsystems up to dual quad-core processor subsystems. Each Cisco WAAS device, regardless of form factor, is configured with some amount of hard disk storage and a compact flash card. The compact flash card is used for boot-time operation and configuration files, whereas the hard disk storage is used for optimization data (including object cache and Data Redundancy Elimination [DRE]), swap space, software image storage repository, and guest operating system storage in the case of WAVE devices. Having a compact flash card enables the device to remain accessible on the network should the device suffer hard drive subsystem failure for troubleshooting and diagnostics purposes

(in such a scenario, optimization and virtualization services would not be operational). Also, by using the compact flash card in this way, a WAAS device can successfully boot and become accessible on the network if no disks are available to the device.

The foundational layer of the Cisco WAAS software is the underlying Cisco Linux platform. The Cisco Linux platform is hardened to ensure that rogue services are not installed and secured such that third-party software or other changes cannot be made. The Cisco Linux platform hosts a command-line interface (CLI) shell similar to that of Cisco IOS Software, which, along with the Central Manager and other interfaces, form the primary means of configuring, managing, and troubleshooting a device or system. All relevant configuration, management, monitoring, and troubleshooting subsystems are made accessible directly through this CLI as opposed to exposing the Linux shell.

The Cisco Linux platform hosts a variety of services for WAAS run-time operation. These include disk encryption, Central Management Subsystem (CMS), interface manager, reporting facilities, network interception and bypass, application traffic policy (ATP) engine, and kernel-integrated virtualization services, as shown in Figure 2-1.

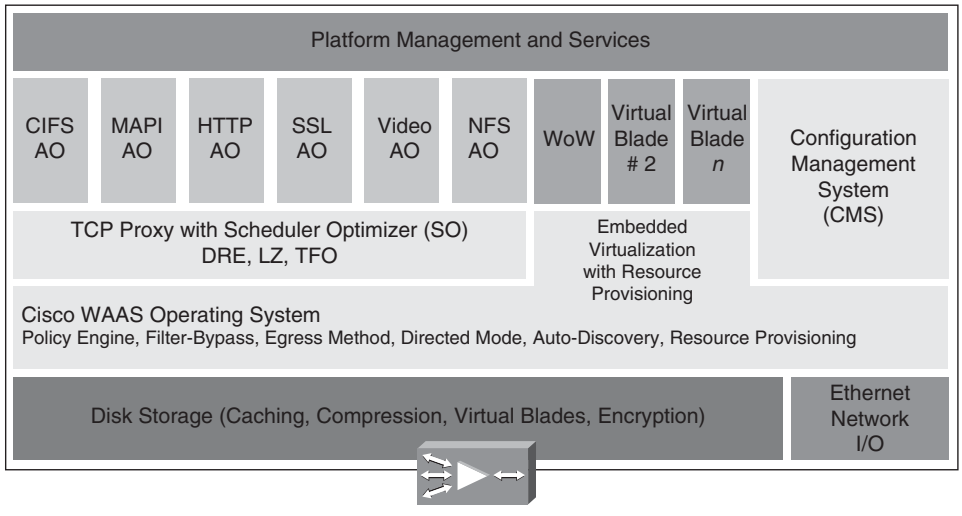

Figure 2-1 *Cisco WAAS Hardware and Software Architecture*

The following sections examine each of the Cisco WAAS architecture items. Cisco WAAS optimization components, including Data Redundancy Elimination (DRE), Persistent LZ Compression (PLZ), Transport Flow Optimization (TFO), and application accelerators, are discussed in detail in Chapter 1, and thus are not discussed in this chapter.

Disk Encryption

Cisco WAAS devices can be configured to encrypt the data, swap, and spool partitions on the hard disk drives using encryption keys that are stored on and retrieved from the Central Manager. The disk encryption feature uses AES-256 encryption, the strongest

commercially available encryption, and keys are stored only in the WAAS device memory after they have been retrieved from the Central Manager during the device boot process. Should a WAAS device be physically compromised or a disk stolen, power is removed from the device, which destroys the copy of the key in memory (memory is not persistent). When the hard disks are encrypted, loss of the key renders data on the disk unusable and scrambled. Keys are stored in the Central Manager database (which can be encrypted) and synchronized among all Central Manager devices for high availability. If a WAAS device is not able to retrieve its key from the Central Manager during boot time, it remains in pass-through mode until connectivity is restored or disk encryption is administratively bypassed. Additionally, the fetching of the key from the Central Manager is done over the Secure Sockets Layer (SSL)-encrypted session that is used for message exchanges between the WAAS devices and the Central Manager devices.

Central Management Subsystem

CMS is a process that runs on each WAAS device, including accelerators and Central Managers. This process manages the configuration and monitoring components of a WAAS device and ensures that each WAAS device is synchronized with the Central Manager based on a scheduler known as the Local Central Manager (LCM) cycle. The LCM cycle is responsible for synchronizing the Central Manager CMS process with the remote WAAS device CMS process to exchange configuration data, fetch health and status information, and gather monitoring and reporting data. The CMS process is tied to a management interface configured on the WAAS device known as the primary interface, which is configured on the WAAS device CLI prior to registration to the Central Manager. Any communication that occurs between WAAS devices for CMS purposes is done using SSL-encrypted connections for security.

Interface Manager

The Cisco WAAS device interface manager manages the physical and logical interfaces that are available on the WAAS device. Each WAAS device includes two integrated Gigabit Ethernet interfaces (including the network modules, one interface is internal and shares connectivity to a peer interface in the router through the router backplane, the other is external and can be cabled to a LAN switch, similar to an appliance). Each WAAS appliance has expansion slots to support one or more additional feature cards, such as the inline bypass adapter, which has two two-port fail-to-wire pairs. The interface manager also provides management over logical interfaces that can be configured over physical interfaces. Logical interfaces include active/standby interfaces, where one physical interface is used as a primary interface and a second interface is used as a backup in the event the primary interface fails. Another logical interface is the PortChannel interface, which can be used to team WAAS device interfaces together for the purposes of high availability and load balancing. It should be noted that active/standby interfaces are used when WAAS device interfaces connect to separate switches, whereas PortChannel interfaces are used when the WAAS device interfaces connect to the same switch.

Monitoring Facilities and Alarms

Cisco Linux provides an interface for the Cisco WAAS software to use for purposes of monitoring and generating alarms. Cisco WAAS supports the Simple Network Management Protocol (SNMP) versions 1, 2c, and 3, and a host of Management Information Bases (MIB) that provide complete coverage over the health of each individual WAAS device. Cisco WAAS also supports the definition of up to four syslog servers, which can be used as alarm recipients when syslog messages are generated. The WAAS Central Manager also has an alarm dashboard, which is described in Chapter 7, "System and Device Management." The Central Manager makes an application programming interface (API) available for third-party visibility systems, which is also discussed in Chapter 7, Chapter 8, "Configuring WAN Optimization," and Chapter 9, "Configuring Application Acceleration." Transaction logs can be configured to be stored on each of the accelerator devices in the network for persistent retention of connection statistics, which might be useful for troubleshooting, debugging, or analytics purposes. Transaction logs are not covered in this book, but a full reference on their usage can be found in the Cisco WAAS documentation.

Note The alarm book (which covers syslog messages, SNMP traps, and Central Manager dashboard alarms), error book (which covers console messages), and product documentation can be downloaded from Cisco.com at http://www.cisco.com/cgi-bin/tablebuild.pl/waas41.

Network Interception and Bypass Manager

The network interception and bypass manager is used by the Cisco WAAS device to establish relationships with intercepting devices where necessary and ensure low-latency bypass of traffic that the WAAS device is not intended to handle. The Web Cache Coordination Protocol version 2 (WCCPv2) is a protocol managed by the network interception and bypass manager to allow the WAAS device to successfully join a WCCPv2 service group with one or more adjacent routers, switches, or other WCCPv2-capable server devices. WCCPv2 is discussed in more detail in Chapter 4, "Network Integration and Interception." Other network interception options, which are also discussed in Chapter 4, include policy-based routing (PBR), physical inline interception, and Application Control Engine (ACE). As flows are intercepted by the WAAS device and determined to be candidates for optimization, those flows are handed to the Application Traffic Policy (ATP) engine to identify what level of optimization and acceleration should be applied based on the configured policies and classifier matches. The ATP is discussed in the next section, and Chapter 8 and Chapter 9 discuss the configuration and management of policies.

Application Traffic Policy Engine

Although the foundational platform component of Cisco WAAS is Cisco Linux, the foundational optimization layer of the Cisco WAAS software (which is as much a component of the Cisco Linux platform as it is the software) is the ATP engine. The ATP is responsible for examining details of each incoming flow (after being handled by the interception and bypass mechanisms) in an attempt to identify the application or protocol associated with the flow. This association is done by comparing the packet headers from each flow against a set of predefined, administratively configured, or dynamic classifiers, each with its own set of one or more match conditions. Flows that do not have a match with an existing classifier are considered "other" traffic and are handled according to the policy defined for other traffic, which indicates that there are no classifier matches and that the default policy should be used.

When a classifier match is found, the ATP examines the policy configuration for that classifier to determine how to optimize the flow. The ATP also notes the application group to which the classifier belongs to route statistics gathered to the appropriate application group for proper charting (visualization) and reporting. The configured policy dictates which optimization and acceleration components are enacted upon the flow and how the packets within the flow are handled. The list of configurable elements within a policy include the following:

- **Type of policy:** Defines whether the policy is a basic policy (optimize, accelerate, and apply a marking), Wide Area File Services Software (WAFS) transport (used for legacy mode compatibility with WAAS version 4.0 devices), and end-point mapper (EPM, used to identify universally-unique identifiers for classification and policy).
- **Application:** Defines which application group the statistics should be collected into, including byte counts, compression ratios, and others, which are then accessible via the WAAS device CLI or Central Manager.
- **Action:** Defines the WAN optimization policy that should be applied to flows that match the classifier match conditions. This includes:
  - **Passthrough:** Take no optimization action on this flow
  - **TFO Only:** Apply only TCP optimization to this flow, but no compression or data deduplication
  - **TFO with LZ Compression:** Apply TCP optimization to this flow, in conjunction with persistent LZ compression
  - **TFO with Data Redundancy Elimination:** Apply TCP optimization to this flow, in conjunction with data deduplication
  - **Full Optimization:** Apply TCP optimization, persistent LZ compression, and data duplication to this flow
- **Accelerate:** Accelerate the traffic from within this flow using one of the available application accelerators. This provides additional performance improvement above

and beyond those provided by the WAN optimization components defined in Action and includes (the capabilities are described in detail in Chapter 1):

- **MS Port Mapper:** Identify application based on its universally unique identifier, which allows WAAS to appropriately classify certain applications that use server-assigned dynamic port numbers
- **Common Internet File System (CIFS):** Acceleration for Microsoft file-sharing environments
- **HTTP:** Acceleration for intranet and Internet applications that use the hypertext transfer protocol
- **NFS:** Acceleration for UNIX file-sharing environments
- **MAPI:** Acceleration for Microsoft Exchange e-mail, calendaring, and collaboration environments
- **Video:** Acceleration for Windows Media over RTSP streams

- **Position:** Specify the priority order of this policy. Policies are evaluated in priority order, and the first classifier and policy match determines the action taken against the flow and where the statistics for that flow are aggregated.
- **Differentiated Services Code Point (DSCP) Marking:** Apply a DSCP value to the packets in the flow. WAAS can either preserve the existing DSCP markings or apply a specific marking to the packets matching the flow based on the configuration of this setting.

Settings configured in the policy are employed in conjunction with one another. For instance, the CIFS policy is, by default, configured to leverage the CIFS accelerator prior to leveraging the "full optimization" (DRE, PLZ, TFO) capabilities of the underlying WAN optimization layer. This can be coupled with a configuration that applies a specific DSCP marking to the packets within the flow. This is defined in a single policy, thereby simplifying overall system policy management. Classifiers within the ATP can be defined based on source or destination IP addresses or ranges, TCP port numbers or ranges, or universally-unique identifiers (UUID). The ATP is consulted only during the establishment of a new connection, which is identified through the presence of the TCP synchronize (SYN) flag which occurs within the first packet of the connection. By making a comparison against the ATP using the SYN packet of the connection being established, the ATP does not need to be consulted for traffic flowing in the reverse direction, as the context of the flow is established by all WAAS devices in the path between the two endpoints and applied to all future packets associated with that particular flow. In this way, classification performed by the ATP is done once against the three-way handshake (SYN, SYN/ACK packets) and is applicable for both directions of traffic flow.

Figure 2-2 shows how the ATP engine interacts with a flow and a particular policy. For more information on ATP, including configuration, please see Chapter 8 and Chapter 9.

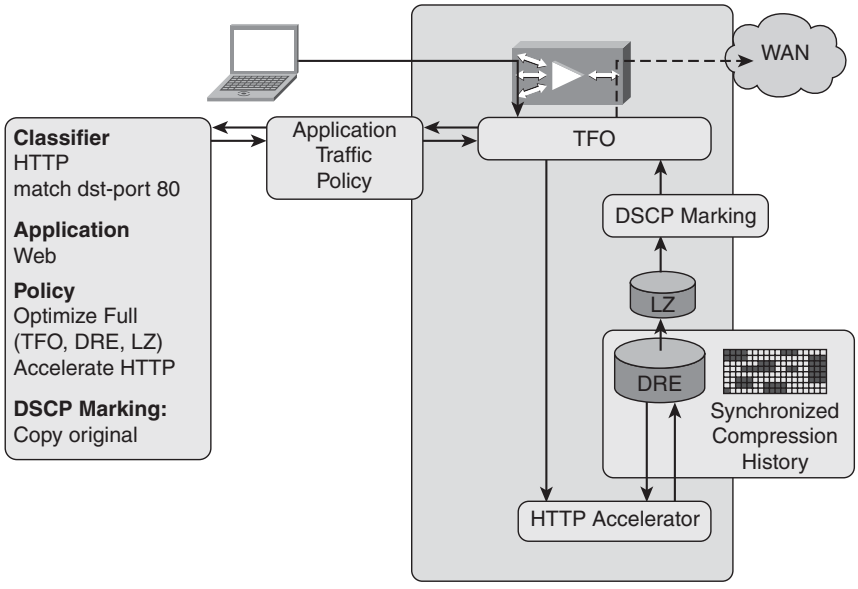

Figure 2-2 *Connection Interaction with Application Traffic Policy*

Virtual Blades

Cisco WAAS utilizes Kernel-based Virtual Machine (KVM) technology from Red Hat (via the Qumranet acquisition) to allow the WAVE appliance (and the WAE-674) to host third-party operating systems and applications. As of version 4.1.3, Microsoft Windows Server, versions 2003 and 2008, are supported for installation on the WAAS Virtual Blade (VB) architecture, and certain configurations can be bundled and packaged within the WAVE configuration with full support from the Cisco Technical Assistance Center (TAC). This configuration includes Microsoft Windows Server 2008 Core, Active Directory read-only domain controller, DNS server, DHCP server, and print server. The WAAS VB architecture helps enable customers to further consolidate infrastructure by minimizing the number of physical servers required in the branch office for those applications which are not good candidates for centralization into a data center location.

Hardware Family

The current Cisco WAAS hardware family consists of three router-integrated network modules, two desktop appliance models, and four rack-mounted appliance models. With such a diverse hardware portfolio, Cisco WAAS can be deployed in each location with the appropriate amount of optimization capacity for the needs of the users or servers in that particular location. This section examines the specifics of each of the current and legacy hardware platforms and positioning of each. Performance and scalability metrics for each are examined later in this chapter, along with best practices around accurately sizing a Cisco WAAS deployment.

Router-Integrated Network Modules

The Cisco WAAS router-integrated network modules are designed to provide optimization services for the remote branch office or enterprise edge. These modules, which are single-processor systems based on the Network Module Enhanced (NME) hardware, can occupy an empty or available NME-capable slot in a Cisco Integrated Services Router (ISR), including models 2811, 2821, 2851, 3825, and 3845. The ISR is an ideal platform for the branch office in that it provides a converged service platform for the remote office, including routing, switching, wireless, voice, security, and WAN optimization in a single chassis (platform, software version, and slot capacity dependent). In addition, the ISR provides a strong foundation for application performance management (APM) solutions in that along with WAAS, other performance-related features can be configured, including quality of service (QoS) for network provisioning, Performance Routing (PfR) for optimal path selection and network utilization, and NetFlow for visibility into traffic distribution, throughput, and other metrics.

Figure 2-3 shows a picture of the Cisco NME-WAE family of WAAS integrated network modules and the ISR family.

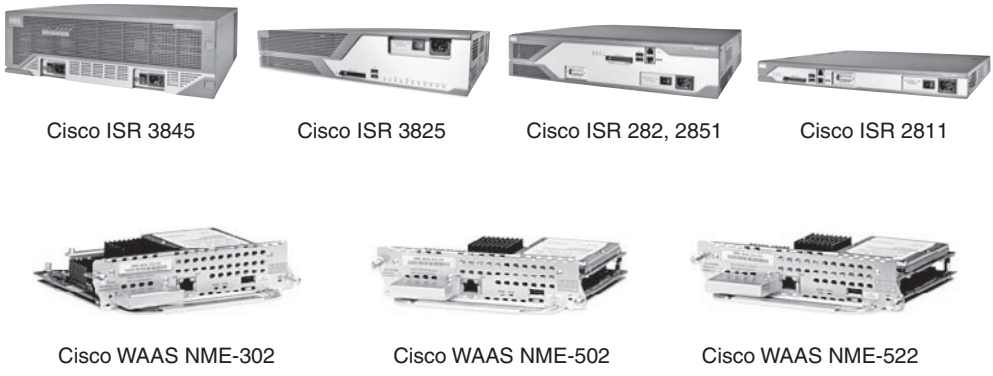

Figure 2-3 *Cisco ISR Family and WAAS Network Modules*

The Cisco NME-WAE family includes three models: the NME-WAE-302, NME-WAE-502, and NME-WAE-522. Each network module has a single hard disk with capacity ranging from 80 to 160 GB. With only a single drive, the NME-WAE is not capable of Redundant Array of Inexpensive Disks (RAID). NME-WAE devices integrate into the network using WCCPv2 as a means of interception (Policy-Based Routing [PBR] can also be used, but WCCPv2 is preferred). Both methods of integration and interception are discussed in Chapter 4. The NME-WAE family does not provide support for virtualization in the branch office; a WAVE appliance model or WAE-674 is required for virtualization support. Each NME-WAE has two network interfaces:

- **One internal:** Connected to the ISR backplane, which communicates with an internal network interface on the ISR

- **One external:** Accessible through the front of the module, which can be attached to a LAN switch

Figure 2-4 shows the architecture of the NME, internal and external interfaces, and intersection points between the NME and the ISR.

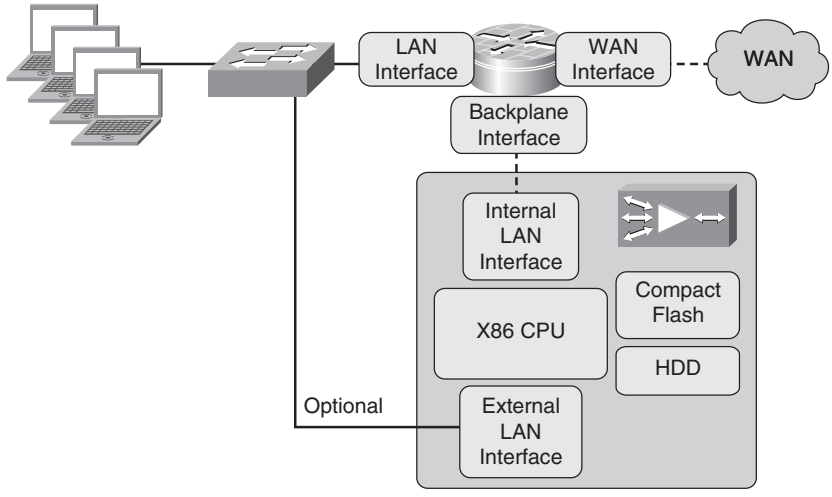

Figure 2-4 *Cisco WAAS Network Module Architecture*

NME-WAE Model 302

The Cisco NME-WAE model 302 (NME-WAE-302) is designed for customers who want to employ only basic WAN optimization capabilities, which are permitted through the use of the Transport license (licensing is discussed later in this chapter). These capabilities include the ATP engine, DRE, PLZ, and TFO. This module is not capable of running the advanced services enabled by the Enterprise license (discussed later in the chapter), including application layer acceleration or disk encryption. The NME-WAE-302 is a single-processor system with 512 MB of RAM and a single 80-GB hard disk.

NME-WAE Model 502

The Cisco NME-WAE model 502 (NME-WAE-502) is designed for customers who want to employ WAN optimization capabilities and application acceleration features for an enterprise edge location. The NME-WAE-502 can be configured with the Enterprise license, providing full WAN optimization functionality, application acceleration functionality, and other features enabled by the Enterprise license including disk encryption and NetQoS integration. The NME-WAE-502 is a single-processor system with 1 GB of RAM and a single 120-GB hard disk. The NME-WAE-502 is capable of supporting a larger number of users than the NME-WAE-302, as discussed in the "Performance and Scalability Metrics" section later in this chapter.

NME-WAE Model 522

The Cisco NME-WAE model 522 (NME-WAE-522) is designed for customers who want to employ appliance-equivalent functionality to an enterprise edge location in the ISR. The NME-WAE-522 supports the full suite of Enterprise license features, including all WAN optimization and application acceleration capabilities. The NME-WAE-522 is a single-processor system with 2 GB of RAM and a 160-GB hard disk, serving as the most powerful network module available as of this writing.

Appliances

The Cisco WAAS appliance family is designed to be deployed in a location of any size, including the small branch office, campus networks, or the largest of enterprise data center networks. The Cisco WAAS appliance family includes the WAE and the newer WAVE devices. Cisco WAVE appliances are current-generation and provide support for branch office virtualization, whereas WAE appliances (with the exception of the WAE-674) do not. The Cisco WAE family includes models 512, 612, 674, 7341, and 7371, and the Cisco WAVE family includes models 274, 474, and 574. WAE appliance models 512 and 674, along with WAVE appliance models 274, 474, and 574, are targeted toward branch office deployments, whereas the WAE appliance models 674, 7341, 7371 are targeted toward regional office and data center deployments. The WAE-674 is a hybrid device that is commonly used for larger branch offices (and those where virtualization is required), but works equally well as a data center device where virtualization is not used. This should not imply that the device characterization presented is fixed; devices should be placed in locations according to performance and scalability sizing and feature requirements.

The WAE appliance models 512, 612, 674, 7341, and 7371, along with WAVE appliance model 574, each have externally accessible hard disk drives and RAID support (some models support hot-swappable disk drives). WAVE appliance models 274 and 474 do not have externally accessible hard disk drives, and with a single hard disk drive, do not support RAID.

Each WAE and WAVE appliance has two built-in Gigabit Ethernet interfaces, which can be deployed independently of one another or as a pair in either an active/standby configuration or PortChannel configuration. Such interface configurations are discussed in Chapter 5, “Branch Office Network Integration,” and Chapter 6, “Data Center Network Integration.” The WAE and WAVE appliance families both have one or more Peripheral Component Interconnect (PCI) expansion slots that support installation of additional feature cards, such as the physical in-path interception card. Each WAE or WAVE appliance can be deployed using a variety of network interception techniques, including physical inline interception, WCCPv2, PBR, and ACE (all are described in Chapter 4). Any appliance model can be used as a core (data center) or edge (branch office) device, although performance and scalability recommendations presented in this chapter must be followed. Figure 2-5 shows an image of the Cisco WAE appliance family, and Figure 2-6 shows an image of the Cisco WAVE appliance family.

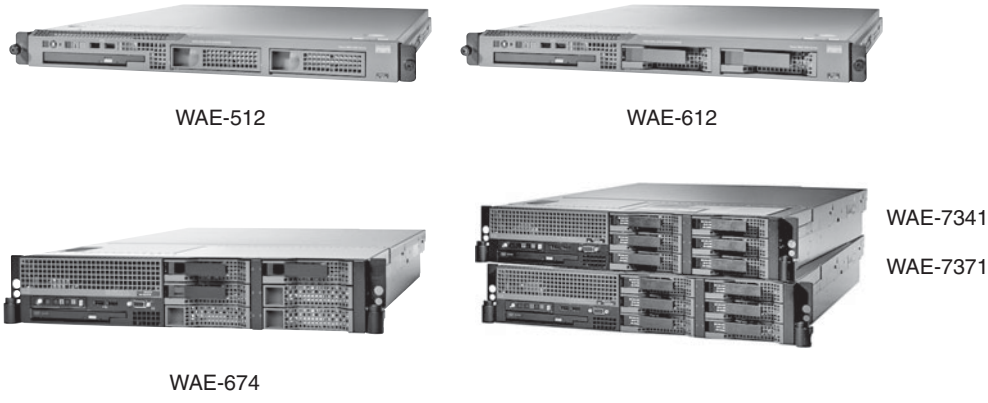

Figure 2-5 *Cisco WAAS WAE Appliance Family*

Figure 2-6 *Cisco WAAS WAVE Appliance Family*

Note The WAE model 7326 is end-of-life and is not covered in this section; however, its performance and scalability metrics are covered in this chapter to support those who have already deployed these devices in their networks and wish to continue using them with WAAS v4.1.

WAVE Model 274

The Cisco WAVE model 274 (WAVE-274) is a single-processor desktop model that is designed for deployment in small and medium-sized branch office locations or small data center locations. The WAVE-274 is configured with 3 GB of RAM. The WAVE-274 provides full WAN optimization and application acceleration capabilities and supports virtualization with up to two VBs. The WAVE-274 can be configured with any license available for WAAS. The WAVE-274 includes a single 250-GB SATA2 hard disk drive, and therefore does not support RAID. The WAVE-274 includes an inline card (with support for one WAN link) and the Enterprise license (discussed in the "Licensing" section of this chapter).

WAVE Model 474

The Cisco WAVE model 474 (WAVE-474) is a single-processor desktop model that is designed for deployment in small- and medium-sized branch office locations or small data center locations. Like the WAVE-274, the WAVE-474 is configured with 3 GB of

RAM. The WAVE-474 provides full WAN optimization and application acceleration capabilities and supports virtualization with up to two VBs. The WAVE-474 can be configured with any license available for WAAS. The WAVE-474 includes a single 250-GB SATA2 hard disk drive, and with a single drive, it does not support RAID. The WAVE-474 is similar to the WAVE-274, but supports a larger number of optimized TCP connections and higher levels of WAN bandwidth. The WAVE-474 includes an inline card (with support for two WAN links) and the Enterprise license (discussed in the "Licensing" section of this chapter).

WAE Model 512

The Cisco WAE model 512 (WAE-512) is a single-processor rack-mount system that is designed for deployment in small- and medium-sized branch office locations or small data center locations. The WAE-512 can be configured with 1 or 2 GB of RAM. In either configuration, the WAE-512 can provide full WAN optimization and application acceleration capabilities, but does not support virtualization. With an increase in memory configuration, the WAE-512 supports a larger number of optimized TCP connections and a greater amount of WAN bandwidth. Regardless of memory configuration, the WAE-512 can be configured with the Transport, Enterprise, or Video license. The WAE-512 supports two 250-GB SATA2 hard disk drives, which are configured automatically for software RAID-1.

WAVE Model 574

The Cisco WAVE model 574 (WAVE-574) is a quad-core rack-mount system that is designed for deployment in large branch office locations or small data center locations. The WAVE-574 can be configured with either 3 GB or 6 GB of RAM and either one or two 500 GB SATA hard disk drives. With two drives, the system is configured automatically for software RAID-1. The 6 GB RAM configuration affords the WAVE-574 support for increased WAN bandwidth and optimized TCP connections and enables the 574 to increase its VB support from two to six (assuming 512MB of RAM is allocated for each VB). The WAVE-574 supports the full breadth of features and capabilities offered by any available Cisco WAAS license.

WAE Model 612

The Cisco WAE model 612 (WAE-612) is a dual-core processor rack-mount system that is designed for deployment in medium-sized branch office locations or medium-sized data center locations. The WAE-612 can be configured with 2 GB or 4 GB of RAM (4 GB of RAM provides greater WAN bandwidth support and higher optimized TCP connection counts) and, in any configuration, supports the full breadth of features and capabilities offered by the Transport, Enterprise, and Video licenses. The WAE-612 supports two 300-GB SAS hard disk drives, which are configured automatically for software RAID-1 and are hot-swap capable.

WAE Model 674

The Cisco WAE model 674 (WAE-674) is a quad-core rack-mount system that is designed for deployment in large branch office locations or medium to large data center locations. The WAE-674 can be configured with either 4 GB or 8 GB of RAM and three 300 GB SAS hard disk drives, which are capable of hot-swap. The 4 GB RAM configuration affords the WAE-674 support for up to two VBs, and the 8 GB RAM configuration affords the WAE-674 support for up to six VBs. Additionally, the increased memory configuration provides support for a greater amount of WAN bandwidth and optimized TCP connections. The WAE-674 is unique in that it can be configured with or without VB support (the only device in the hardware family that can), and when configured without VB support, the WAE-674 can support an even higher level of WAN bandwidth and optimized TCP connections. The reason for this level of configurability is the unique position of the WAE-674, which can be used for branch offices and data centers of virtually any size. The WAE-674 supports the full breadth of features and capabilities offered by any available Cisco WAAS license.

WAE Model 7341

The Cisco Cisco WAE model 7341 (WAE-7341) is a single quad-core rack-mount system (four processors) that is designed for deployment in large enterprise data centers. The WAE-7341 includes 12 GB of RAM and four 300-GB Serial-Attached SCSI (SAS) hard disk drives, which are configured automatically for hardware RAID-5 and support hot-swap. The WAE-7341 supports the full breadth of features and capabilities offered by the Transport, Enterprise, and Video Cisco WAAS licenses, but not virtualization.

WAE Model 7371

The Cisco WAE model 7371 (WAE-7371) is a dual quad-core rack-mount system (eight processors) that is designed for deployment in the largest of enterprise data centers and under the most demanding conditions. The WAE-7371 includes 24 GB of RAM and six 300-GB SAS hard disk drives, which are configured automatically for hardware RAID-5 and support hot-swap. The WAE-7371 supports the full breadth of features and capabilities offered by the Transport, Enterprise, and Video Cisco WAAS licenses, but not virtualization.

Licensing

Each Cisco WAAS device, whether it is an appliance (WAE or WAVE) or a router-integrated network module, must be configured with one or more licenses. This license dictates what features are permitted to be configured on the device. Licenses are not enforced in WAAS; however, licenses can only be applied to platforms that support the

particular license in question. Four licenses exist for Cisco WAAS and configuration of licenses are discussed in Chapter 7:

- **Transport license:** Enables a WAAS device to apply only basic WAN optimization capabilities. It supports use of TFO, DRE, and PLZ. WAAS devices configured with the Transport license cannot provide Enterprise license features including application-acceleration capabilities, disk encryption, or any other features provided by other licenses. WAAS devices configured with the Transport license can, however, register with and be managed and monitored by a WAAS device configured as a Central Manager. The Transport license is supported by all Cisco WAAS hardware platforms.
- **Enterprise license:** Allows a WAAS device to apply all the WAN optimization provided by the Transport license and all the application acceleration functionality with the exception of Video (which is licensed separately). Additionally, the Enterprise license enables support for disk encryption and NetQoS integration. Like the Transport license, WAAS devices configured with the Enterprise license can register with and be managed and monitored by a WAAS device configured as a Central Manager. Configuration of a WAAS device as a Central Manager requires the Enterprise license. The Enterprise license is supported by all Cisco WAAS hardware platforms with the exception of the network module model 302 (NME-302).
- **Video:** Allows a WAAS device to apply stream splitting to Windows Media over Real-Time Streaming Protocol (RTSP) traffic. The Video license is commonly applied in conjunction with the Enterprise license. The Video license is supported by all Cisco WAAS hardware platforms with the exception of the network module model 302 (NME-302).
- **Virtual-Blade:** Allows a WAAS device to host third-party operating systems and applications in one or more VBs in the branch office, including Microsoft Windows Server. The Virtual-Blade license is supported on all Cisco WAVE appliances in addition to the WAE model 674.

Performance and Scalability Metrics

Design of a Cisco WAAS solution involves many factors, but the cornerstone of the solution design is based on the performance and scalability metrics required for the solution as a whole and for each individual location where WAAS is deployed. Every component in an end-to-end system has a series of static and dynamic system limits. For instance, a typical application server might be limited in terms of the number of connections it can support, disk I/O throughput, network throughput, CPU speed, or number of transactions per second. Likewise, each Cisco WAAS device has static and dynamic system limits that dictate how and when a particular WAAS device is selected for a location within an end-to-end design. This section examines the performance and scalability metrics of the Cisco WAAS hardware family, and provides a definition of what each item is and how it is relevant to a localized (per location) design and an end-to-end system design.

The static and dynamic limits referred to are used as a means of identifying which device is best suited to provide services to a particular location in the network. The

device might be deployed as an edge device, where it connects to potentially many peer devices in one or more data center locations, or as a core device, where it serves as an aggregation point for many connected edges. WAAS devices can also be deployed as devices to optimize links between data center locations, where devices on each side are realistically core devices. A fundamental understanding of the performance and scalability metrics is paramount in ensuring a sound design. Although WAAS devices have no concept of "core" or "edge," the deployment position within the network has an effect on the type of workload handled by a device and should be considered—primarily as it relates to TCP connection count and peer fan-out (how many peers can connect to a device for the purposes of optimization). This section examines each of the performance and scalability system limits, both static and dynamic, that should be considered. These include device memory, disk capacity, the number of optimized TCP connections, WAN bandwidth and LAN throughput, the number of peers and fan-out, and the number of devices managed.

Device Memory

The amount of memory installed in a device dictates the level of performance and scalability the device can provide. As the memory capacity increases, the ability of a WAAS device to handle a larger number of connections, a larger addressable index space for compression, or a longer history of compression data also increases. Having larger amounts of memory also enables the WAAS device to run additional services, such as application acceleration, disk encryption, or virtualization, and positions the device to accept additional features that might be introduced in future software releases.

The NME-WAE family members have fixed memory capacity and cannot be upgraded. Thus, the system limits for the NME-WAE family are static. From the WAE appliance family, the 7341 and 7371 have fixed memory configurations. However, the WAE-512, WAE-612, and WAE-674 have configurable memory options, in that:

- The WAE-512 can be configured with 1 GB or 2 GB of memory.
- The WAE-612 can be configured with 2 GB or 4 GB of memory.
- The WAE-674 can be configured with 4 GB or 8 GB of memory.

For devices that support flexible memory configuration (such as the WAE-512, WAE-612, and WAE-674), higher levels of WAN bandwidth can be realized, along with an increase in the number of optimized TCP connections that can be handled concurrently by that device. For virtualization-capable platforms, a larger number of VBs can be supported. The WAVE appliance family models 274 and 474, like the network modules, are fixed configuration and do not support a memory upgrade, whereas the 574 model—like the WAE 512, 612, and 674—does support memory configuration (either 3 GB or 6 GB).

The amount of installed memory directly impacts what license is supported on each of the device models. The Transport license can be configured on any WAAS hardware model. WAAS hardware models that have 1 GB of memory or more (all do except the NME-WAE-302) can be configured with the Enterprise license, which allows the WAAS device to operate all of the Enterprise license features.

Previous versions of Cisco WAAS (version 4.0.x and version 4.1.x when using legacy mode compatibility) had distinct *core* and *edge* CIFS acceleration services. With legacy mode, a device with 1 GB of RAM can support only edge services for CIFS, whereas a device with 2 GB of RAM or more can support edge or core services, or both together. As of Cisco WAAS version 4.1.1, this deployment mode is no longer required unless interoperability with version 4.0.x is required. Generally speaking, most customers upgrade the entire network in a short and well-defined period of time and can take advantage of the simplified deployment model provided in 4.1.x, which does not have such restrictions.

Disk Capacity

Optimization services in the Cisco WAAS hardware family leverage both memory and disk. From a disk perspective, the larger the amount of available capacity, the larger the amount of optimization history that can be leveraged by the WAAS device during runtime operation. For instance, an NME-WAE-502 has 120 GB of physical disk capacity, of which 35 GB is available for use by DRE for compression history. With 35 GB of compression history, one can estimate the length of the compression history given WAN conditions, expected network utilization, and assumed redundancy levels.

Table 2-1 shows how the length of the compression history can be calculated for a particular WAAS device, along with an example. This example assumes a T1 WAN that is

Table 2-1 *Calculating Compression History*

Step	Action	Example Result
1	Convert WAN capacity to bytes (divide the number of bits per second by 8)	(T1 = 1.544 Mbps) / 8 = 193 KBps
2	Identify maximum WAN throughput for a given day (convert from seconds to minutes, to hours, to a single day)	193 KB/sec * 60 sec/min 11.58 MB/min * 60 min/hr 694.8 MB/hr * 24 hr/day Total 16.68 GB/day
3	Identify WAN throughput given utilization (multiply by the number of hours and utilization per hour)	(694.8 MB/hr * 8 hours) * 75% utilization = 4.168 GB (694.8 MB/hr * 16 hours) * 50% utilization = 5.56 GB Total = 9.72 GB/day
4	Identify WAN throughput given utilization and expected redundancy (multiply daily throughput by expected redundancy or compressibility)	9.72 GB/day * .25 (as .75 is 75% redundancy) = 2.43 GB/day
5	Calculate compression history (divide capacity by daily throughput)	Storage capacity of unit divided by daily throughput 35 GB / 2.43 GB/day = 14.4 days of history

75 percent utilized during business hours (75 percent utilization over 8 hours per day) and 50 percent utilized during nonbusiness hours (16 hours per day), and assumes that data traversing the network is 75 percent redundant (highly compressible by DRE). This table also assumes an NME-WAE-502 with 35 GB of allocated capacity for DRE compression history.

It is generally recommended that, at minimum, five days of compression history be available in a WAAS device to better ensure that substantial performance improvements are possible. In the example in Table 2-1, the NME-WAE-502 contains enough storage capacity to provide an effective compression history of two weeks. In most cases, users tend to access data that is newer more frequently, whereas older data is accessed less frequently. Because of this, having five days worth of compression history could even be considered overkill.

The disk capacity available to a WAAS device is split among five major components:

- **DRE compression history:** This capacity is used for storing DRE chunk data and signatures.
- **CIFS cache:** This capacity is preallocated on all devices using the Enterprise license.
- **Print services:** This capacity is preallocated for print spool capacity. Print services require that the Enterprise license be configured and that CIFS edge services be configured, which implies that legacy mode is being used. In cases where print services are configured, the 1 GB of disk capacity is allocated. Given that 1 GB is a fraction of the total storage capacity of a device, it is not accounted for in Table 2-2.
- **Platform services:** This capacity is preallocated for operating system image storage, log files, and swap space.
- **Virtual Blades:** This capacity is preallocated for any guest operating systems and applications that are installed to run in a WAAS VB.

Table 2-2 shows the storage allocation for each WAAS device for each of these components.

Number of Optimized TCP Connections

Each WAAS device has a static number of TCP connections that can be optimized concurrently. Each TCP connection is allocated memory and other resources within the system, and if the concurrently optimized TCP connection static limit is met, additional connections are handled in a pass-through fashion. Adaptive buffering (memory allocation) is used to ensure that more active connections are allocated additional memory, and less active connections are only allocated the memory they require.

The TCP connection limit of each WAAS device can be roughly correlated to the number of users supported by a given WAAS device model, but note that the number of TCP connections open on a particular node can vary based on user productivity, application behavior, time of day, and other factors. It is commonly assumed that a user will have 5 to

Table 2-2 *Disk Capacity Allocation per Platform*

Platform	Total Usable Capacity	DRE	CIFS	VBs
NME-WAE-302	80 GB	30 GB	0 GB	0 GB
NME-WAE-502	120 GB	35 GB	49 GB	0 GB
NME-WAE-522	160 GB	67 GB	67 GB	0 GB
WAVE-274	250 GB	40 GB	120 GB	35 GB
WAVE-474	250 GB	60 GB	120 GB	35 GB
WAE-512-1GB	250 GB RAID-1	60 GB	120 GB	0 GB
WAE-512-2GB	250 GB RAID-1	80 GB	100 GB	0 GB
WAVE-574-3GB	500 GB RAID-1	80 GB	120 GB	60 GB
WAVE-574-6GB	500 GB RAID-1	120 GB	120 GB	180 GB
WAE-612-2GB	300 GB RAID-1	100 GB	120 GB	0 GB
WAE-612-4GB	300 GB RAID-1	120 GB	120 GB	0 GB
WAE-674-4GB	600 GB RAID-5	120 GB	120 GB	120 GB
WAE-674-8GB	600 GB RAID-5	150 GB (with VB) 320 GB (without VB)	120 GB	200 GB (with VB) 0 GB (without VB)
WAE-7326	900 GB RAID-1	320 GB	230 GB	0 GB
WAE-7341	900 GB RAID-5	500 GB	230 GB	0 GB
WAE-7371	1500 GB RAID-5	1 TB	230 GB	0 GB

15 connections open at any given time, with roughly 6 to 10 of those connections requiring optimization. If necessary, policies can be adjusted on the WAAS Central Manager to pass through certain applications that might realize only a small amount of benefit from WAAS. This type of change could potentially help increase the number of users that can be supported by a particular WAAS device.

Table 2-3 shows the optimized TCP connection capacity per device model.

Table 2-3 *Optimized TCP Connection Capacity per Platform*

Network Module	Connection Capacity	Appliance	Connection Capacity
NME-WAE-302	250	WAVE-274	200
NME-WAE-502	500	WAVE-474	400
NME-WAE-522	800	WAE-512-1GB	600
		WAE-512-2GB	1200
		WAVE-574-3GB	750
		WAVE-574-6GB	1300
		WAE-612-2GB	1600
		WAE-612-4GB	4800
		WAE-674-4GB	2000
		WAE-674-8GB (with VB)	4000
		WAE-674-8GB (without VB)	6000
		WAE-7326	5000
		WAE-7341	12,000
		WAE-7371	50,000

The number of connections a typical user has in a location can be determined by using tools that exist in the operating system of the user's workstation. Although the estimate of six to ten optimized TCP connections is accurate for the broad majority of customers, those that wish to more accurately determine exactly how many connections a typical user has open at any given time can do so.

Microsoft provides two methods for determining the number of connections that are open on a given computer. The first is through the Command Prompt program **netstat**. By opening a Command Prompt window (click **Start > Run**, then type **cmd** and click **Ok**) and typing the command **netstat**, you can see a list of the open connections from the computer to all of the other endpoints to which that computer is connected. Notice the connections that are in the state of ESTABLISHED. These connections are currently open and in use and have not yet been closed. In many cases, the protocol associated with the connection is listed next to the foreign address, but some might not be. From here, you can identify the servers to which the user is connected and determine which should and should not be optimized. Figure 2-7 shows an example of the output of this command.

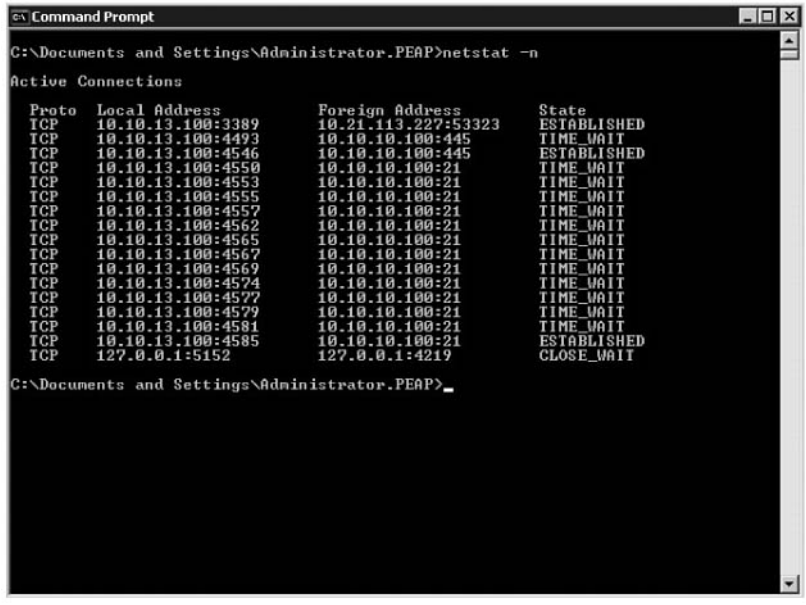

Figure 2-7 *Determining the Number of TCP Connections In Use Using netstat*

Another tool provided by Microsoft that (along with many other things) provides visibility into the number of TCP connections in use on a particular computer is Performance Monitor. Performance Monitor can be accessed by clicking **Start > Run** and typing **perfmon**, followed by clicking **Ok**. From within the Performance Monitor window, click the **+** sign, select the TCP performance object, and then add the **Connections Established** counter. Doing so shows you the number of connections established over time, and this data can even be exported for offline use. Figure 2-8 illustrates an example output from Performance Monitor showing the number of established TCP connections.

Linux, UNIX, and Macintosh provide similar tools to understand the number of connections that are open on a given computer. The **netstat** command is available on virtually any Linux distribution and is available in most UNIX platforms and versions of Apple's Macintosh OS/X operating system.

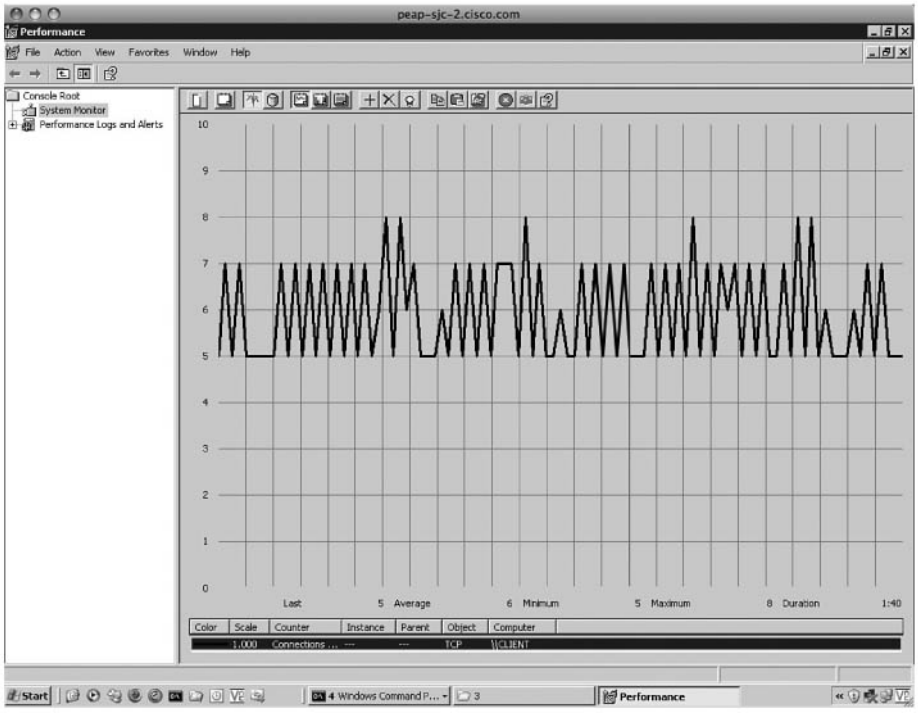

Figure 2-8 *Determining the Number of TCP Connections in Use Using Performance Monitor*

For the data center, the sum of all remote office TCP connections should be considered one of the key benchmarks by which the data center sizing should be done. Note that the largest Cisco WAAS device supports up to 50,000 optimized TCP connections—which is approximately 5,000 users (assuming ten TCP connections per user). For organizations that need to support a larger number of users or want to deploy the data center devices in a high-availability manner, multiple devices can be used. The type of network interception used (discussed in Chapter 4) determines the aggregate number of optimized TCP connections that can be supported by a group of Cisco WAAS devices deployed at a common place within the data center. Recommended practice dictates that sites that require high availability be designed with *N+1* availability in consideration relative to the number of maximum optimized TCP connections—that is, if 100,000 optimized TCP connections must be supported, the location should have a minimum of two WAE-7371 devices to support the workload, a third WAE-7371 device to handle failure of one of the devices, and use an interception mechanism such as WCCP or ACE that supports load-balancing of workload across the entire set of three devices. Other considerations apply, as discussed in Chapter 4.

WAN Bandwidth and LAN Throughput

WAAS devices are not restricted in software or hardware in terms of the amount of WAN bandwidth or LAN throughput supported. However, recommendations are in place to specify which WAAS device should be considered for a specific WAN environment. WAN bandwidth is defined as the amount of WAN capacity that the WAAS device can fully use when employing the full suite of optimization capabilities (this includes DRE, PLZ, TFO, and the other application acceleration capabilities). LAN throughput is defined as the maximum amount of application layer throughput (throughput as perceived by the users and servers) that can be achieved with the particular WAAS hardware model and an equivalent or more-powerful peer deployed at the opposite end of the network.

For some deployment scenarios, it is desired to use the Cisco WAAS devices only for TCP optimization. Cisco WAAS TFO provides a powerful suite of optimizations to better allow communicating nodes to "fill the pipe" (that is, fully leverage the available WAN bandwidth capacity) when the application protocol is not restricting throughput due to application-induced latency. Each Cisco WAAS device has a TFO-only throughput capacity that can be considered when WAAS devices are deployed strictly for TCP optimization only. This is recommended only for situations where compression, redundancy elimination, and application acceleration are not required, and the application throughput has been validated to be hindered only by the performance of the TCP implementation in use. This is common in some data center to data center applications—such as data replication or data protection—where the traffic that is sent is previously compressed, redundancy eliminated, or encrypted. TFO attempts to fully utilize the available bandwidth capacity, but might be hindered by congestion in the network (not enough available bandwidth) or performance impedance caused by application protocol chatter.

Table 2-4 shows the WAN bandwidth supported by each WAAS device model and the maximum LAN-side throughput and TFO-only throughput capacity. Note that other factors can influence these values and throughput levels can be achieved only when the link capacity available supports such a throughput level. For instance, a LAN throughput maximum of 150 Mbps is not possible on a Fast Ethernet connection; rather, a Gigabit Ethernet connection is required. Similarly for throughput speeds more than 1 Gbps, multiple 1-Gbps interfaces must be used.

The amount of bandwidth required per site is the sum of available WAN capacity that can be used at that site and not the sum of all WAN bandwidth for every connected peer. For instance, if a branch office has four bundled T1 links (totaling 6 Mbps of aggregate WAN throughput) but only two are used at any given time (high availability configuration), a device that supports 3 Mbps or more is sufficient to support the location.

Table 2-4 *WAN Bandwidth and LAN Throughput Capacity per WAAS Device*

WAAS Device Model	WAN Supported	LAN Throughput Maximum	TFO-Only Throughput Maximum
NME-WAE-302	4 Mbps	90 Mbps	100 Mbps
NME-WAE-502	4 Mbps	150 Mbps	150 Mbps
NME-WAE-522	8 Mbps	2000 Mbps	250 Mbps
WAVE-274	2 Mbps	90 Mbps	150 Mbps
WAVE-474	4 Mbps	90 Mbps	250 Mbps
WAE-512-1GB	8 Mbps	100 Mbps	350 Mbps
WAE-512-2GB	20 Mbps	150 Mbps	400 Mbps
WAVE-574-3GB	8 Mbps	100 Mbps	350 Mbps
WAVE-574-6GB	20 Mbps	150 Mbps	400 Mbps
WAE-612-2GB	45 Mbps	250 Mbps	450 Mbps
WAE-612-4GB	90 Mbps	350 Mbps	500 Mbps
WAVE-674-4GB	45 Mbps	250 Mbps	450 Mbps
WAVE-674-8GB (with or without VB)	90 Mbps	350 Mbps	500 Mbps
WAE-7326	155 Mbps	450 Mbps	600 Mbps
WAE-7341	310 Mbps	800 Mbps	800 Mbps
WAE-7371	1 Gbps	1.5 Gbps	1.8 Gbps

Similarly, if a data center has four DS-3 links (totaling 180 Mbps of aggregate WAN throughput) but uses only three at a time (*N+1* configuration), a device that supports 135 Mbps of WAN bandwidth or more is sufficient to support that location. The WAN throughput figures mentioned in the preceding table are (as discussed previously) not limited in hardware or software. In some cases, the WAN throughput that a device achieves might be higher than the values specified here. Those interested in using a smaller device to support a larger WAN link (for instance, qualifying a WAVE-274 for an 8-Mbps ADSL connection) are encouraged to test the system under those conditions and validate the performance prior to making a decision to use that specific platform.

Number of Peers and Fan-Out

Each Cisco WAAS device has a static system limit in terms of the number of concurrent peers it can actively communicate with at any one given time. When designing for a particular location where the number of peers exceeds the maximum capacity of an individual device, multiple devices can be deployed, assuming an interception mechanism that uses load balancing is employed (such as WCCPv2 or ACE; these are discussed in

Chapter 4). In cases where load balancing is used, TCP connections are distributed according to the interception configuration, thereby allowing for near-linear scalability increases in connection count, peer count, and WAN bandwidth, as devices are added to the pool. Load-balancing interception techniques are recommended when multiple devices are used in a location, and in general, an *N+1* design is recommended.

Peer relationships are established between Cisco WAAS devices during the automatic discovery process on the first connection optimized between the two devices. These peer relationships time out after ten minutes of inactivity (that is, no active connections are established and optimized between two peers for ten minutes). Each WAAS device supports a finite number of active peers, and when the peer relationship is timed out, that frees up peering capacity that can be reused by another peer. Data stored in the DRE compression history remains intact even if a peer becomes disconnected due to inactivity, unless the DRE compression history becomes full. In cases where the DRE compression history becomes full, an eviction process is initiated to remove the oldest set of data in the DRE compression history to make room for new data.

Table 2-5 shows the maximum number of concurrent peers supported per WAAS platform. If peers are connected beyond the allocated limit, the WAE permits the connections to be

Table 2-5 *Maximum Supported Peers per WAAS Device*

Network Module	Concurrent Peers	Appliance	Recommended Concurrent Peers
302	5	WAVE-274	35
502	15	WAVE-474	35
522	40	512-1GB	35
		512-2GB	70
		WAVE-574-3GB	35
		WAVE-574-6GB	70
		612-2GB	210
		612-4GB	350
		WAVE-674-4GB	100
		WAVE-674-8GB (no VB)	200
		WAVE-674-8GB (with VB)	200
		7326	600
		7341	1400
		7371	2800

established and gracefully degrades performance as needed. Connections associated with peers in excess of the maximum fan-out ratio are able to use the existing compression history but are not able to add new chunks of data to it. The end result is lower effective compression ratios for the connections using peers that are in excess of the specified fan-out ratio.

The number of peers supported by a device is typically the last factor that should be considered when sizing a solution for a particular location. The primary reason being that the WAN capacity or number of connections supported at the maximum concurrent peers specification is generally an order of magnitude higher than what the device can support. For instance, although a WAE-7371 can support up to 2800 peers, even if those peers were the NME-302 (each supporting 250 optimized TCP connections), it is not able to handle the 700,000 possible optimized TCP connections that all 2,800 NME-302s were attempting to optimize with it. It is best to size a location first based on WAN bandwidth capacity and TCP connections, and in most cases, only a simple validation that the number of peers supported is actually required.

Number of Devices Managed

Each Cisco WAAS deployment must have at least one Cisco WAAS device deployed as a Central Manager. The Central Manager is responsible for system-wide policy definition, synchronization of configuration, device monitoring, alarming, and reporting. The Central Manager can be deployed only on appliances and can be deployed in an active/standby fashion. When a certain WAAS device is configured as a Central Manager, it is able to, based on the hardware platform selected for the Central Manager, manage a maximum number of WAAS devices within the topology. Only WAAS appliances can be configured as Central Manager devices, and in high-availability configurations, each Central Manager WAE should be of the same hardware configuration. Although hardware disparity between Central Manager WAEs works, it is not a recommended practice given the difference in the number of devices that can be managed among the WAE hardware models. It should be noted that standby Central Managers (such a configuration is examined in Chapter 7) receive information in a synchronized manner identical to how accelerator WAAS devices do. Table 2-6 shows the maximum number of managed nodes that can be supported by each WAAS appliance when configured as a Central Manager.

Use of multiple WAAS devices configured as Central Manager devices do not increase the overall scalability in terms of the number of devices that can be managed. To manage a number of devices greater than the capacities mentioned in the preceding table, multiple autonomous Central Managers are needed. For instance, in an environment with 3000 devices, two separate instances of Central Manager are required, and each instance can be comprised of a single device or multiple devices deployed in a high availability primary/standby configuration.

Table 2-6 *Central Manager Scalability*

Appliance	Managed Nodes
WAVE-274	125
WAVE-474	250
WAE-512-1GB	500
WAE-512-2GB	750
WAVE-574-3GB	500
WAVE-574-6GB	1000
WAE-612-2GB	750
WAE-612-4GB	1500
WAE-674-4GB	1500
WAE-674-8GB	2000

Replication Acceleration

The WAE-7341 and WAE-7371 devices support a deployment mode called *Replication Accelerator*, which requires Cisco WAAS version 4.0.19, or a version newer than that from the 4.0 train. This mode of acceleration is used for data center to data center deployments where replication and backup acceleration is required, and when configured, adjusts the behavior of the WAAS device to allocate larger blocks of memory to a smaller number of connections, and minimizes the processing latency of DRE by using only memory for deduplication. Although only memory is used for DRE, the DRE data is persistent in that it is written to disk, but the disk is used only to reload the previous compression history. This enables WAAS to provide high levels of throughput necessary to accelerate replication and backup traffic between data centers.

The network typically found in these cases is high-bandwidth and relatively low latency (above 10–20 ms), where a significant amount of data needs to be moved from one location to another location in a short period of time. The performance and scalability metrics of replication accelerator mode are different than the performance and scalability metrics that would normally be considered for these devices when not deployed in replication accelerator mode and are documented in Table 2-7.

Table 2-7 *Replication Accelerator Performance and Scalability Metrics*

Appliance	WAN Bandwidth	LAN Throughput	Optimized TCP Connections	Concurrent Peers	DRE Capacity
WAE-7341	310 Mbps	800 Mbps	2500	4	12 GB
WAE-7371	1 Gbps	1.5 Gbps	5000	9	24 GB

Although all WAAS devices in a given network can be managed by a common Central Manager, WAAS devices configured in replication accelerator mode can only peer with other WAAS devices that are configured as replicator accelerator devices. Should intermediary application accelerator devices exist in the network path between two replication accelerator devices (this is generally rare, as replication accelerator devices are deployed between backend networks as opposed to the enterprise WAN), the application accelerator devices are not able to peer with replication accelerator devices.

Replication accelerator devices are commonly deployed on backend data center to data center networks and not the enterprise WAN due to the high bandwidth requirements. WAAS devices configured as replication accelerators are commonly found deployed as follows:

- **Directly attached to one or more storage array IP/Ethernet interfaces:** Such a deployment model dedicates the devices to optimize replication for that particular array and that particular interface.
- **Directly attached to one or more storage fabric switch or director IP/Ethernet interfaces:** Including the Cisco MDS 9000 family, such a deployment model enables the devices to optimize replication or backup traffic traversing fabrics in distant sites over IP.
- **Directly behind the data center interconnect device:** Such a deployment model enables optimization of any traffic between data centers. In this deployment model, replication accelerator should be carefully considered against the standard application accelerator mode which may be more applicable in cases where a large body of nonreplication and nonbackup traffic exists.

Virtual Blades

The Cisco WAVE appliance family and the WAE-674 provide branch office virtualization capabilities that enable consolidation of remote branch office servers onto the WAAS device as a shared platform. Sizing for VBs should be done in conjunction with sizing for WAN optimization and application acceleration because the available disk capacity to support VBs and the number of VBs supported varies per platform based on the hardware configuration as shown in Table 2-8.

To accurately size a virtualization solution for a branch office, it is necessary to understand the minimum and recommended memory requirements to support the operating system and applications you plan to install on top of that operating system. Many vendors support installation of their server operating system onto systems with only 512 MB of memory, which increases the maximum number of VBs that can be installed on a WAAS device; however, many have requirements for larger amounts of memory.

Additionally, consider the disk capacity requirements necessary for each VB, and reconcile that amount with the total VB storage capacity of the platform selected for that given location. Even the smallest virtualization-capable WAAS device (the WAVE-274) supports 35 GB of disk capacity for VBs—meaning that with two VBs, configured, you have

Table 2-8 *VB Capacity*

Appliance	VB Disk Capacity	VB Memory Capacity	Maximum Number of VBs (512 MB RAM each)
WAVE-274	35 GB	1 GB	2
WAVE-474	35 GB	1 GB	2
WAVE-574-3GB	60 GB	1 GB	2
WAVE-574-6GB	180 GB	3 GB	6
WAE-674-4GB	120 GB	1 GB	2
WAE-674-8GB	200 GB	3 GB	6

approximately 17.5 GB of disk space for each. Storage capacity allocation is flexible in that you can allocate as much space as is available from the pool to any particular VB. However, you should ensure that you size the system for the location with enough capacity to support the current application and operating system requirements as well as future requirements. More information on configuration and deployment of VBs can be found in Chapter 10, "Branch Office Virtualization."

Summary

The Cisco Wide-Area Application Engine family includes three network modules for the Integrated Services Router and six appliance models spanning two desktop models and four rack-mount appliance models. This breadth of portfolio provides customers with the flexibility necessary to allocate the right platform for each network location where WAN optimization, application acceleration, and virtualization capabilities are needed. Four licenses are available for Cisco WAAS, including the Transport license (WAN optimization capabilities only), Enterprise license (all application accelerators except video, and certain other features), Video (Windows Media over RTSP stream splitting), and VirtualBlades (branch office virtualization platform). Sizing of a Cisco WAAS solution requires consideration of a number of factors, including network conditions (WAN bandwidth and LAN throughput), number of users and concurrent optimized TCP connections, disk capacity and compression history, memory, concurrently connected peers, and virtualization requirements. By following the recommended guidelines for performance and scalability, a robust Cisco WAAS design can be realized, thereby allowing administrators to deploy the solution confidently to improve application performance over the WAN while enabling centralization and consolidation of costly infrastructure.

Index

A

B

C

D

E

F

G

H

I

J-K

L

M

N

O

P

Q

R

S

T

U

V

W

X-Z