

Quality of Service White Paper

**“Integrated QoS:
IBM WebSphere and Cisco Can Deliver End-to-End Value”**

e-Business Challenges Driving the Demand for End-to-End Quality of Service (QoS)

The explosive growth of Internet e-business has generated new requirements for the network and computing infrastructure in addition to new business dynamics. As burgeoning new web-based applications stress the IT infrastructure, businesses are looking for a way to prioritize Internet traffic and transactions across the entire infrastructure. Instead of over-provisioning, businesses now need to reduce infrastructure costs while creating new revenue opportunities. More specifically, businesses now must be able to:

- distinguish revenue-generating transactions from routine usage and ensure prioritized end-to-end processing and transport of these transactions
- reduce the number of abandoned transactions due to response time
- provide distinct service to different users based on value
- deliver distinct service based on content accessed
- optimize network and computer processing
- handle flash crowds without a large commitment of computing resource

QoS can address these issues. Today's QoS offerings can be implemented by either the application, server, network router or switches. The inherent problem is that these approaches can not be implemented across all the network, application and middleware elements. Each element assigns its own relative priority for the Internet traffic which makes it challenging to maintain a relative priority across the entire infrastructure. The answer is *layered and dynamic QoS* which can integrate the existing QoS mechanisms available today through the applications, servers, network routers and switches.

To manage this growth, successful e-businesses will be driving end-to-end QoS technology requirements to help them gain more control and management of their entire network and computing resources. These businesses need to maximize network and computing efficiencies and this requires consistent and predictable performance characteristics. A QoS solution encompassing both *application-aware networks* and *network-aware applications* can meet the needs of the growing e-business or service provider.

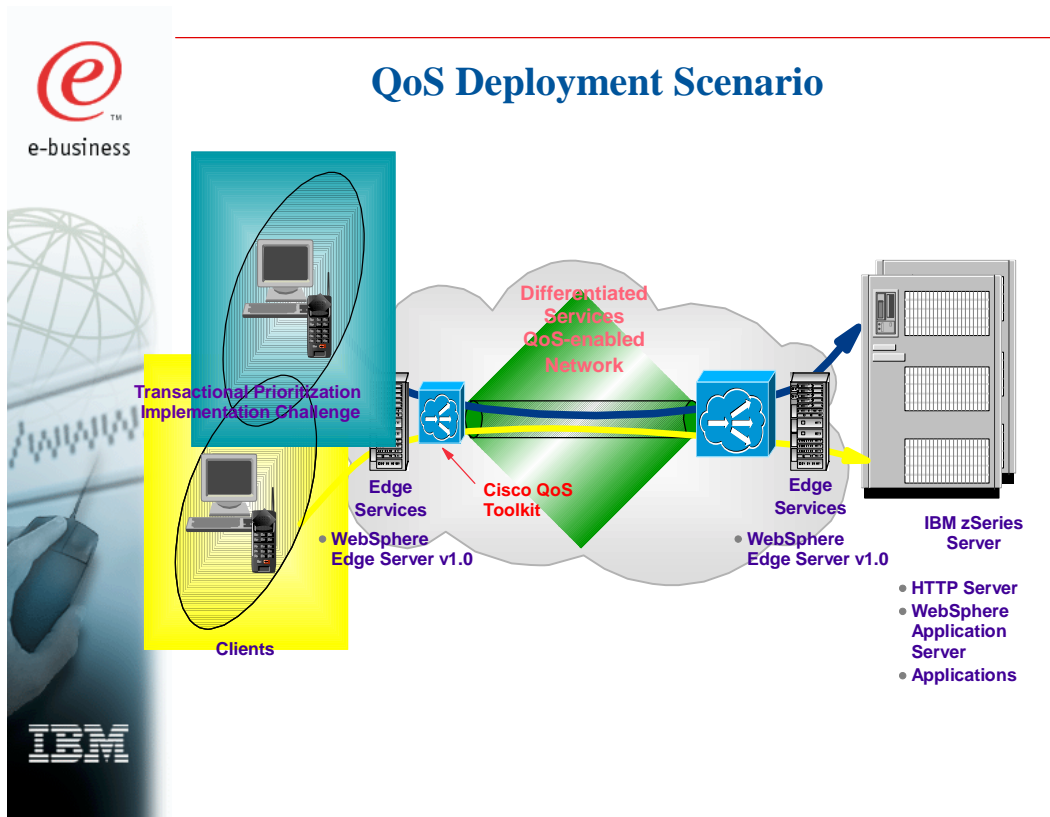
Working together, IBM and Cisco can deliver integrated QoS between IBM WebSphere® applications and middleware and Cisco network routers and switches.

The Need for Integrated Quality of Service

Today's QoS challenges help us understand the value and need for the unique end-to-end QoS solution that IBM WebSphere and Cisco can deliver.

Most QoS approaches today are implemented by the various network, middleware and application elements, independently. Network routers use Type of Service (ToS) fields in packets to set relative priorities for traffic flowing through the network. Software applications and middleware employ mechanisms to prioritize the workload depending on the content-type, processing requirements and available computing resources. The applications and middleware know the kind of network requirements for the various types of traffic that they deliver. These QoS approaches can be effective but independently cannot deliver an integrated QoS solution.

The following chart represents how QoS could be implemented today.



QoS at Network Edges and Middleware

QoS mechanisms at the network edges and middleware typically address server load balancing, content-based routing and distribution, outbound traffic shaping, and inbound traffic/connection control. With an SSL termination point at the network edge by WebSphere Edge Server, encrypted data can be decrypted, which allows for enhanced load balancing. These mechanisms operate closer to the application layer and utilize application specific data (URLs, cookies, content headers etc.) in conjunction with network and server performance metrics and network topology to route requests to appropriate servers to improve the scalability and reliability characteristics of the solution.

QoS for Applications

Applications and servers primarily employ QoS mechanisms to prioritize the allocation of server resources (such as CPU and memory) to specific application tasks. Newer web applications that generate dynamic personalized content have the ability to prioritize a

specific application task (a servlet, for example) based on the state or value of a customer interaction. The state of the art is evolving towards caching of dynamic content, distributed application execution and device specific data transformation.

Network QoS Technology

Differentiated Services (DiffServ) provides service differentiation between broad classes of users and applications. DiffServ aggregates traffic into classes and indicates the different QoS service levels in the IP header. The IETF defines a standard format for Differentiated Services field in the IP packet header in RFC-2474. Examples of queuing and scheduling services used by networking devices to support DiffServ include WFQ (Weighted Fair Queuing), Class-Based Queuing (CBWFQ), WRR (Weighted Round Robin) and WRED (Weighted Random Early Discard).

A full QoS solution has to manage both server capacity and network bandwidth, as interactive applications vie with non-mission-critical data for access to the shared Web infrastructure. The solution is not only ToS but also load balancing, caching, traffic shaping, congestion/connection management, denial of service and security. Again, a QoS solution encompassing both *application-aware networks* and *network-aware applications* can meet the needs of the growing e-business or service provider.

IBM and Cisco are positioned to work together to deliver integrated and interoperable technology to solve this problem.

Turn on the Technology of the End-to-End QoS

There is also a critical need for a policy-based mechanism to assign the relative importance or the definition of business-level objectives which then determine the type of service delivery. Policy mechanisms implement and enforce these objectives using the available and appropriate QoS mechanisms in the network and software infrastructure. Without policy rules, the relative QoS values could not be set. Without end-to-end QoS the business policies could not be implemented consistently. And to achieve true end-to-end QoS there needs to be a well defined mapping between business objectives and the policy mechanisms that apply to network elements.

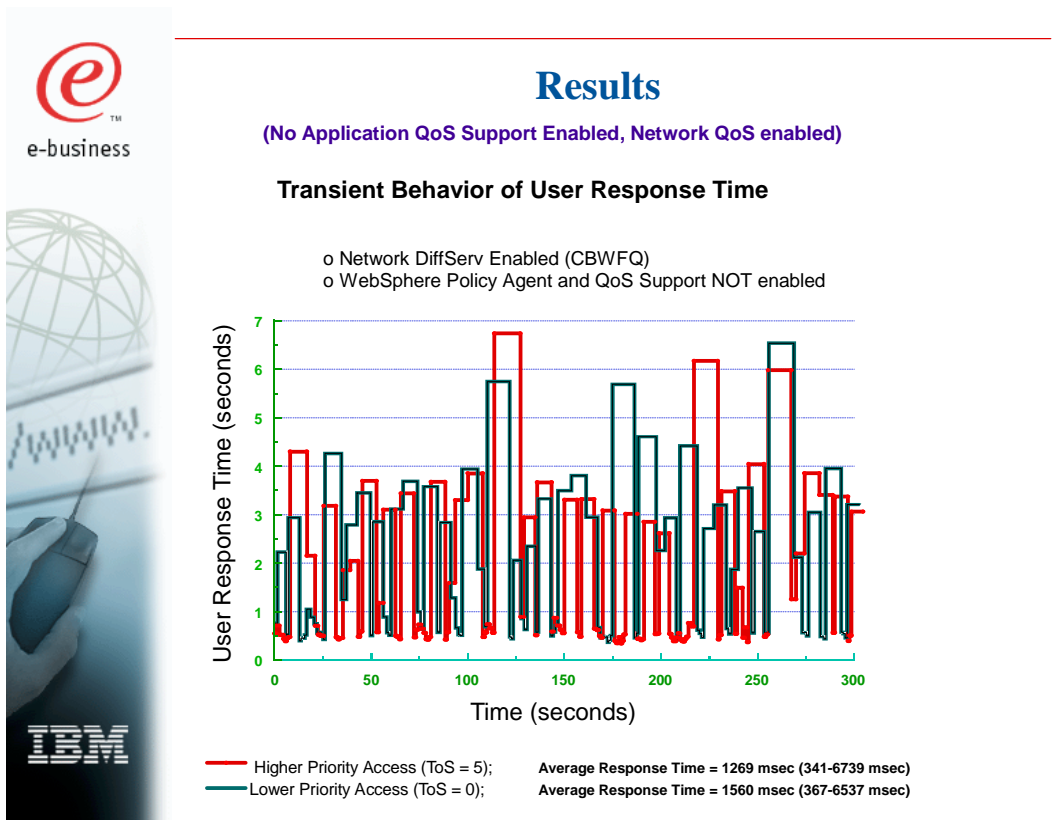
The various application and network QoS mechanisms provide independent views of the relative business priorities for a given transaction. The work load manager (WLM) policies on OS/390® determine appropriate resources allocation for the various application servers. The WebSphere Application Server identifies the type of content based on the URI. WebSphere Application Server then determines the relative priority for that content from the existing OS/390 WLM policy and passes it to the IBM Policy Agent. The IBM Policy Agent determines whether or not there is an existing network QoS policy for the given content. If a network policy does exist, the IBM Policy Agent allows the network QoS classification to take precedence. Otherwise, the IBM Policy Agent uses the application-based QoS classification. The resulting QoS priority is then reflected out into the network. The Cisco network (Cisco IOS software) can perform various queuing algorithms based the QoS classification including the commonly deployed class-based weighted fair queuing methodology.

To help ensure the interoperability of their QoS technologies, IBM and Cisco have worked together to deploy consistent QoS algorithms. As a result, the IBM Policy Agent is capable of effectively mediating between the application-level and network-level policies so that the overall business objectives are implemented end-to-end. For example, CBWFQ employs the appropriate algorithm to enforce the priority set by the IBM Policy Agent.

IBM and Cisco Can Deliver Proven, End-to-End QoS

It is clear that the ability to consistently specify, integrate, enforce and manage the QoS mechanisms available in different parts of an end-to-end solution is critical. The enterprise or service provider need end-to-end QoS to deliver value-added services as well as the ability to manage the solution infrastructure (networking, middleware and servers) to meet SLAs (service level agreements).

Without an end-to-end QoS solution, enterprises and services providers can experience highly variable, unpredictable service delivery. The chart below depicts response time variants found today when using nonintegrated QoS mechanisms on a congested network.



Today, IBM and Cisco are delivering the technology to enable end-to-end QoS that can dramatically improve the response time and variability. Joint testing by IBM and Cisco validated that the IBM S/390 QoS signaling and the Cisco congestion management tools were indeed compatible and that the benefits of consistent response time and effective bandwidth utilization could be realized. More specifically, the product components for this test included the IBM Policy Agent for the Operating System for IBMzSeries Servers (formerly OS/390) for QoS signaling and the Cisco Class-Based Weighted Fair Queuing algorithm in the Cisco IOS software for congestion management. In addition, SSL termination points provided by the WebSphere Edge Server allow for the load balancing of encrypted data for more comprehensive services.

Not all applications can establish priority, and a priority established by an application can conflict with corporate policy. OS/390 provides a Policy Agent that enables the system programmer to overrule or establish priority for any application. As the source and destination of application traffic, a server has full knowledge of all data flows and is therefore effective and efficient in setting QoS service levels. As a result, system administrators can provide QoS signaling at the application source to interface with congestion management technologies in the network. This approach improves network performance eliminating the need to classify the traffic explicitly at each WAN interface in the core or backbone network, helping avoid per-packet processing.

IBM Policy Agent QoS definitions are extensive and can be specified based on application type, individual user, user group, time of day, and day of week. For each policy rule, a corresponding service category defines the appropriate QoS. A service category generally contains the priority of the traffic, the minimum and maximum TCP connection throughput, and the number of TCP connections allowed at any given time. The S/390 can provide this per-connection bandwidth management function through TCP window manipulation. The per-connection level of granularity provided by the S/390 can be used to fine tune the aggregate bandwidth management provided by the router.

The integrity of QoS can only be maintained if these policy rules are applied consistently across the network hardware, middleware and applications. Implementing the QoS classification can be as simple as an application setting the IP precedence bits in all the packets sent. Alternatively, a policy manager running in a server can set IP precedence as appropriate for the applications running in that server, the time of day, the individual end user, or any of a variety of conditions that may be known only to that server—overriding that which was set by the application if necessary. Traffic can also be classified in the network itself, according to network-wide policy based on congestion or multiservice traffic requirements. Regardless of whether the application, server, or network chooses classification, the packet is marked to provide QoS signaling. Without packet marking, every router in the network would have to incur additional overhead to make a policy decision. The policy-based mechanism can allow companies to deliver reliable, manageable end-to-end QoS.

The Business Advantage of IBM and Cisco End-to-End QoS

Businesses need to deliver differentiated e-business services to their customers and partners in response to the increased e-business demand for the same network and application resources. The key to delivering differentiated e-business services is being able to predictably, consistently manage the delivery of Internet traffic and transactions across all the network hardware, application and middleware elements.

The IBM and Cisco integrated QoS technology can deliver proven results. The compelling results from the joint IBM and Cisco testing are best represented by the following chart which depicts performance on a congested network.

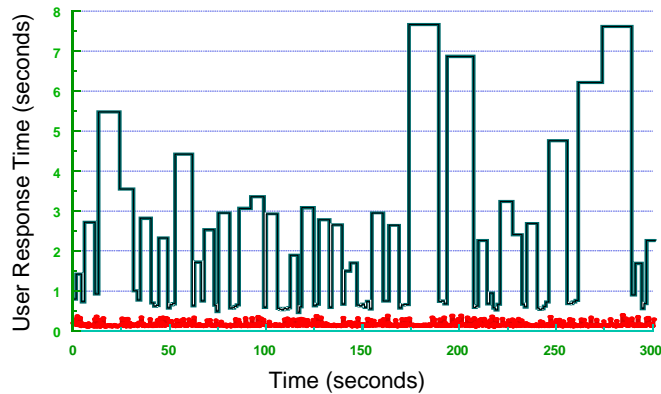


Results

(Application QoS Support Enabled, Network QoS enabled)

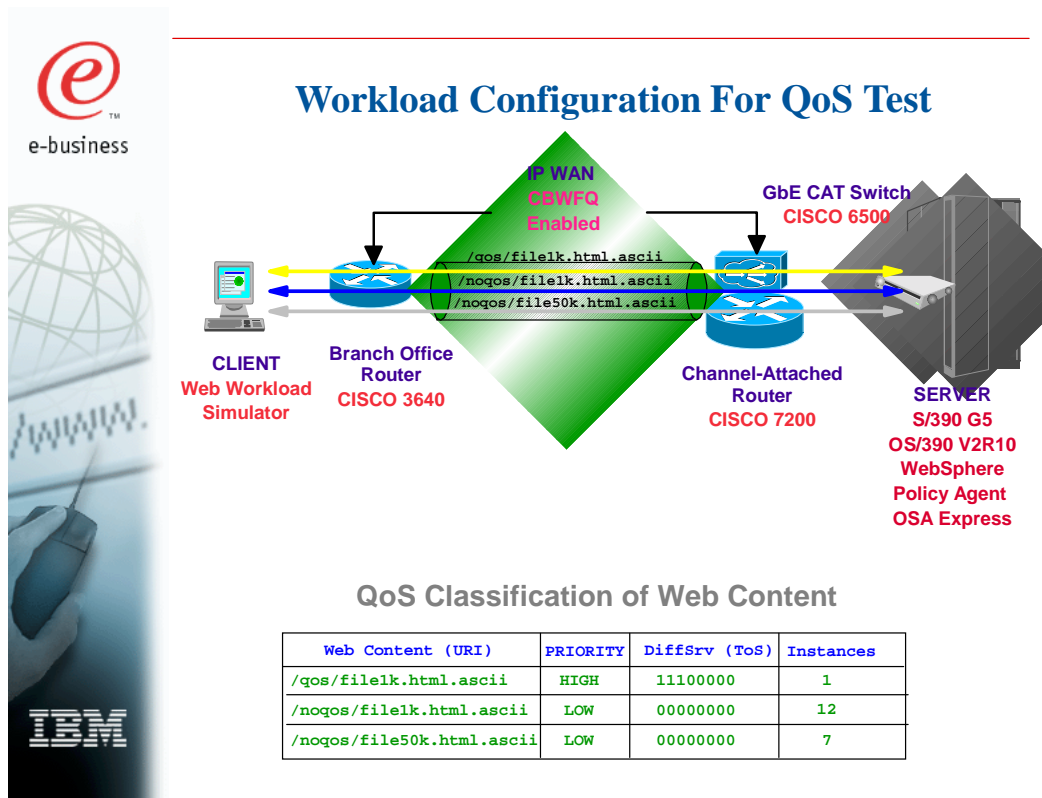
Transient Behavior of User Response Time (WebSphere + IBM HTTP Server for OS/390)

- o Network DiffServ Enabled (CBWFQ)
- o WebSphere Policy Agent and QoS Support Enabled



— Higher Priority Access (ToS = 5); Average Response Time = 157 msec (98-393 msec)
— Lower Priority Access (ToS = 0); Average Response Time = 1506 msec (465-7669 msec)

The testing was conducted using the following test configuration.



With this added control and manageability enabled by the IBM WebSphere and Cisco end-to-end QoS technologies, businesses can now deliver reliable, differentiated or tiered services to their customers. This means that service providers who can offer quality assurances for end-to-end business traffic will can more enterprise business going forward. QoS technologies can allow service providers to offer more services, such as real-time traffic support, or specific bandwidth allocations, to build into an SLA portfolio. And with options such as SSL termination at the edge of network, IBM and Cisco can provide additional flexibility and value. From the IBM and Cisco customer perspective, reliable, manageable tiered services can provide more revenue generation for service providers, while offering more services to enterprises.

IBM®

© Copyright IBM Corporation 2001

IBM Corporation
Department KOJA
3039 Cornwallis Road
Research Triangle Park, NC 27709

Produced in the United States of America
4-01
All Rights Reserved

IBM, WebSphere, WebSphere Edge Server, OS/390, IBM Policy Agent, IBM zSeries, the e-business logo, and OSA Express are trademarks of International Business Machines Corporation in the United States, other countries or both.

Cisco and Cisco IOS are registered trademarks of Cisco Systems, Inc. or its affiliates in the U.S. and certain other countries.

Other company, product and service names may be trademarks or service marks of others.

All statements regarding IBM future direction or intent are subject to change or withdrawal without notice and represent goals and objectives only.