Large-scale High-quality Communication Service Solution Using Active Network Technology

Takashi Nishikado
Minoru Koizumi
Hideo Oochi

OVERVIEW: As more and more people use various Web services from fixed and mobile phones, response times are slowing. Sometimes, users cannot get any response at all because of congestion in communication lines and at Web servers. A new architecture is thus needed to enable quick responses to many users even though communication lines and Web server resources are limited. Prioritized data transfer should also be considered for important users and for special data, such as the latest news and audio/video clips. Hitachi's new communication service solution uses the latest "active network" technology. New-type Web gateway servers are placed at the edge of the network to distribute the processing using cached data following the instructions supplied by Web service providers. For example, providers could instruct the gateway servers on how to prioritize data transfer by using a control tag attached to the contents. Users could then get quicker responses and services differentiated based on the characteristics of themselves and the contents.

INTRODUCTION

NOWADAYS, various services are being provided through the Internet. For example, stocks can be traded, ticket reservations can be made, and advertisement information can be browsed.

In addition, new communication technologies enable Web access from mobile phones. In this environment, people can easily use various Web services from anywhere at anytime.

However, this ease of use causes congestion at networks and Web servers, degrading service quality. Depending on the time and the sites accessed, Web users are frequently forced to wait a long time for a response. This is a serious problem for services providing financial services. Not only is the slow response irritating, users cannot rely on services with poor quality.

Web service providers are working to enhance their infrastructures to provide better quality services. However, it is uneconomical to provide enough communication lines and Web server resources to meet the demand during peak access periods, which can be several times that normally.

Hitachi's new communication service solution uses new-type Web gateway servers that use the latest "active network" technology[1]. Active network technology enables intelligent network operations, that is, data processing in the network as well as data transferring. With this service solution, Web service providers and communication service providers can provide scalable services, and users can get services with higher quality.

In this paper, we first describe the requirements for upcoming Internet services. Then, we explain the new communication service solution provided by Hitachi INS Software, Inc. Finally, we present some service examples using the solution.

REQUIREMENTS FOR UPCOMING INTERNET SERVICES

Of the upcoming Internet services, two types of Web services are expected to become very popular. One type is time-intensive service such as stock trading and ticket reservation. The other type is streaming service such as audio and video data distribution.

As shown in Fig. 1, there are three requirements in particular providing such services with good quality.

Support for Large-scale Services

It is estimated that around a million simultaneous Web accesses are already being made to very frequently accessed Web sites, such as stock trading sites in the U.S. and to the official Olympic game site, at peak times. If such a large number of requests were

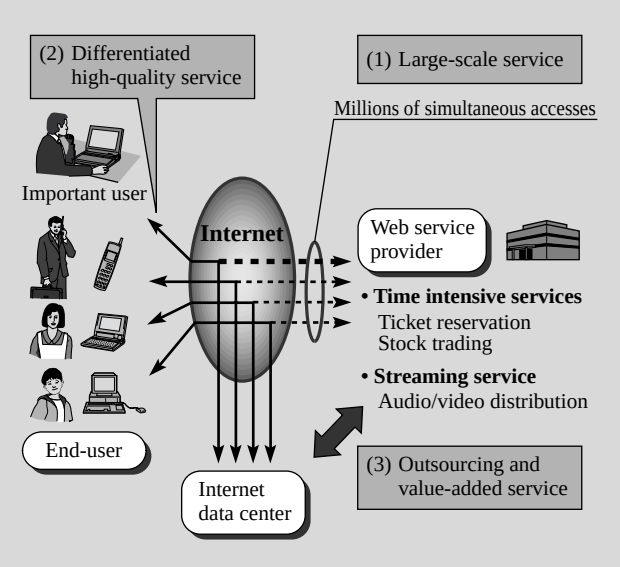

Fig. 1—Requirements for Upcoming Internet Services. Upcoming Internet services will typically have millions of simultaneous accesses. Differentiated, high-quality services macthed to the characteristics of the users and data will be required. Outsourcing and value-added services provided by Internet data centers should also be considered.

processed at a centralized Web site, it would be necessary to provide a line with a bandwidth of 100 Gbit/s–1 Tbit/s and thousands of Web servers at the site in order to provide services with good quality.

Therefore, a new Web communication service solution is necessary to enable quick responses to millions of simultaneous accesses with a limited communication line and a limited number of Web server resource.

Support for Differentiated, High-quality Services

Since communication lines and server resources are limited, important requests may not be served well if all requests are processed equally.

Therefore, communication priority and quality should be controllable based on the users and the data. For example, contents for important users should be processed with higher priority than for regular ones. Special data, such as the latest news and audio/video clips, should be transferred with higher priority and quality.

Consideration of Outsourcing and Value-added Services

Internet data centers that provide outsourcing of Web service have rapidly been increasing in number. The reason for the rapid growth is that Web service providers can reduce their equipment and maintenance costs and provide scaleable and stable services by using the outsourcing services provided by Internet specialists.

Communication service providers are also considering providing value-added services, such as accounting/billing agent services, in addition to conventional communication services.

Therefore, a new mechanism should be considered for such outsourcing and value-added services provided by Internet data centers and communication service providers on behalf of Web service providers.

NEW COMMUNICATION SERVICE SOLUTION USING "ACTIVE NETWORK" TECHNOLOGY

As shown in Fig. 2, our proposed communication service solution satisfies the above requirements.

As mentioned in the Introduction, the active network technology enables intelligent network operations. While the concept of data processing in the network is very attractive, the general approach is still in the research phase. We have thus adjusted the technology to make it practical for use in intelligent Web communication services, and Hitachi INS Software, Inc. has used it to produce a Web gateway (WebGW) server, a software product.

The WebGW server relays Web communication between end-users and Web service providers. By placing these servers at the edge of the network, scalable and high-quality Web services can be provided. The features of the server are as follows.

Content-specified Communication Control

Using the concept of active network, Web service providers can use content data to specify how the data should be treated by the WebGW servers. For example, they can specify access control and communication priority for each data packet. Such communication controls can be implemented using a communication control tag (CCT) attached to each data packet. The tag is written in a standard Web data markup language, XML (extensible markup language)[2].

When a WebGW server receives content with a communication control tag as a reply to an end-user's Web access request, it interprets the tag. It then relays the data to a router, along with the priority information, translated to make it understandable to the router. With this mechanism, providers can prioritize the data transfers to match the characteristics of the users and the contents. This type of fine-grain differentiation

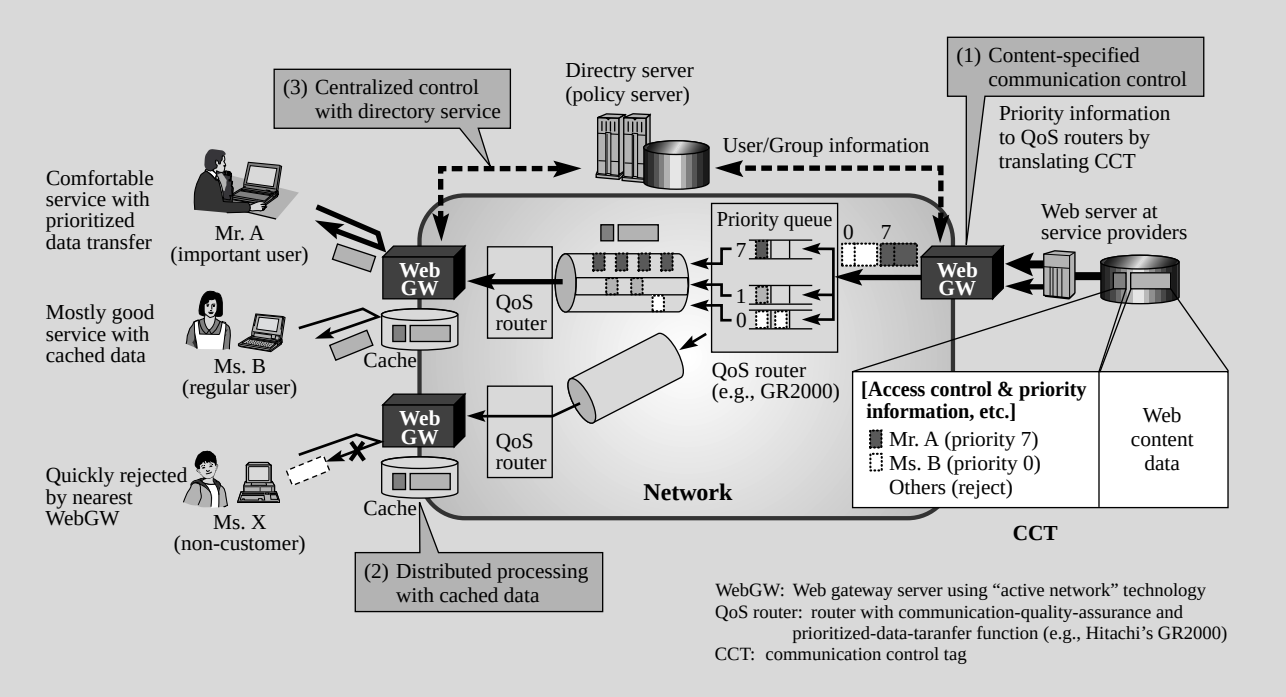

Fig. 2—New Communication Service Solution Based on New-type Web Gateway Servers Using "Active Network" Technology.
Large-scale and high-quality Web services can be provided via distributed Web gateway (WebGW) servers using "active network" technology. Following the instruction of the communication control tag attached to Web content data, WebGW servers can provide differentiated services by controlling QoS routers. They can also support distributed processing using cached data. Central control and management can be provided using the information provided by a directory server.

cannot be provided with only routers, because routers cannot distinguish user and content characteristics.

Additionally, since the communication control is done using a tag attached to the contents, and the tag is removed when sending the contents to the end-user, this service can be provided without changing any of the software of either the end-users or the Web service providers.

Distributed Processing with Cached Data

With WebGW servers, the processing of Web services can be distributed by using cached data. The WebGW server has a caching function to store both content and communication control tags. Since the tags can carry various types of processing information as well as access and QoS control information, various value-added services can be provided by using cached data.

As a result, many requests from users, which would be forwarded to Web servers in a conventional system, can be processed by the nearest WebGW server, if it has the requested contents in cache. As a result, users can get quicker responses even if many users are making requests at the same time.

Centralized Control with Directory Service

The WebGW server can control access and prioritize data transfer for special users and groups by using user/group information and priority service information stored in a directory server.

With this mechanism, access can be controlled and data transfer can be prioritized for Web content without using control tags. If a control tag is supplied, even finer grain service can be provided.

In the future, user/group information and priority service information can be integrated with the policy control information used to control the routers. With this integration, more integrated quality control can be provided with a single policy control server, enabling more dynamic quality control. Using sophisticated traffic measurement technology, providers will be able to flexibly adjust the quality to match network conditions.

SERVICE EXAMPLES FOR INTERNET DATA CENTERS USING NEW COMMUNICATION SERVICE SOLUTION

We showed a service image for a communication service provider in Fig. 2. As other service images, we will present two example services for Internet data centers.

Large-scale Data Distribution Service

Large-scale data distribution service through an Internet data center is illustrated in Fig. 3. In this example, content providers such as game software vendors prepare content data along with licensing information in a communication control tag. The Internet data center interprets the tag and distributes the content to appropriate customers, managing the users and licensing on behalf of the providers. With this service, content providers are not concerned with licensing management and can distribute their contents to a large number of users with simple equipment such as PCs.

Intranet/Extranet Hosting Service

The second example is an intranet/extranet hosting service by an Internet data center (Fig. 4). Because sites in a large enterprise are usually connected via low bandwidth lines, requests from many users cannot be handled, if one or more users start a big data transfer. In this new hosting service, a WebGW server is placed at each gateway of a site, providing prioritized data transfer for important business processings. This makes it possible to smoothly handle requests from important users and for important data at anytime.

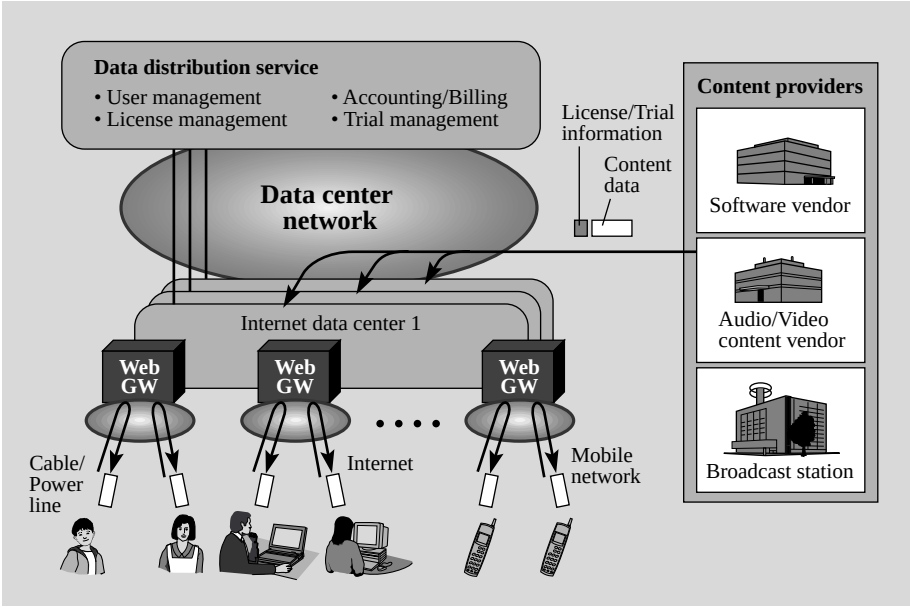

Fig. 3—Large-scale Data Distribution Service. An Internet data center provides large-scale data distribution services with licensing management, etc. on behalf of content providers.

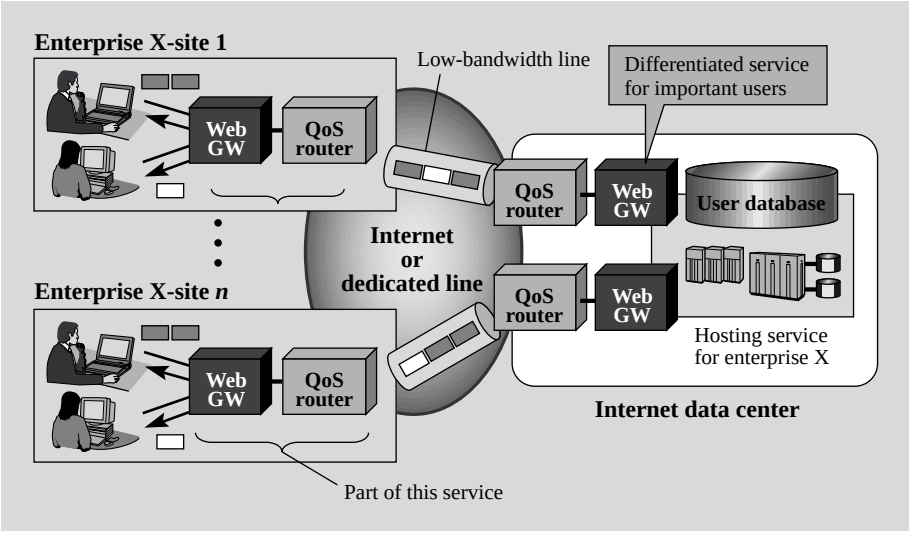

Fig. 4—Intranet/Extranet Hosting Service. Using the prioritized data-transfer function provided by WebGW servers, important business processings hosted in an Internet data center are handled smoothly.

CONCLUSIONS

Hitachi's new communication service solution expands the capacity of time-intensive and streaming services such as ticket reservation and audio/video data distribution, respectively. This solution enables distributed processing by using data cached in the network and the latest active network technology, so quick responses can be provided even if many users are making requests at the same time. Moreover, Web service providers can provide differentiated services by attaching a communication control tag to the contents. Data transfers can be prioritized depending on the characteristics of the users and the contents.

We are going to support other example value-added services, such as accounting/billing agent services, using this solution. We are also going to integrate the management with routers and provide more dynamic quality control functions, ones that adjust based on the current network load.

REFERENCES

(1) D. Wetherall et al., "Introducing New Internet Services: Why and How," IEEE Network, Special Issues on Active and Programmable Networks (July 1998).

(2) W3C, "Extensible Markup Language (XML)," W3C Recommendation (Feb. 1998), http://www.w3.org/TR/REC-xml.

ABOUT THE AUTHORS

Takashi Nishikado

Joined Hitachi, Ltd. in 1984, and now works in Network SE/SI Research Department at the IP Network Research Center. He is currently engaged in the research and the development of information communication systems. Mr. Nishikado is a member of the Information Processing Society of Japan, and can be reached by e-mail at tak@sdl.hitachi.co.jp.

Minoru Koizumi

Joined Hitachi, Ltd. in 1981, and now works in Network SE/SI Research Department at the IP Network Research Center. He is currently engaged in the management of the research and development of management and control technology for network systems. Mr. Koizumi is a member of the Institute of Electronics, Information and Communication Engineers of Japan, and the Information Processing Society of Japan, and can be reached by e-mail at m-koizu@sdl.hitachi.co.jp.

Hideo Oochi

Joined Hitachi, Ltd. in 1972, and now works in the Cyber-Business Division of Hitachi INS Software, Inc. He is currently engaged in the management of the development of Web systems, and can be reached by e-mail at oochi@ins-hitachi.co.jp.