QoS Management System for Video Distribution Services

Akihiko Tsuno*†, *Yoshihiro Otsuka, and Sohei Majima

Abstract

We describe a quality-of-service (QoS) management system for the providers of video distribution services over IP (Internet protocol) networks. It detects degradation in user-perceived QoS by active measurement and locates the cause and it evaluates the QoS perceived by each user by passive measurement. The system consists of active agents, passive agents, traffic generators, HMI (human-machine interface) terminals, management servers, and quality report servers. In this paper, we explain how to detect degradation by active measurement and locate its cause. We also describe an example of QoS evaluation by active and passive measurement.

1. Introduction

Recent developments in the areas of information and communication technologies have been truly remarkable, and the Internet in particular has become almost indispensable to most of us in our lives. The growth of the Internet has certainly hastened the proliferation of broadband access lines and the emergence of an always-on connection environment. Capitalizing on these developments, a growing number of providers have entered the market to deliver video over IP (Internet protocol) networks, with the result that video distribution services have become fairly commonplace. However, we have also seen that video distribution over broadband is vulnerable to quality-of-service (QoS) degradation, so service providers need to manage the quality of the elements of the content distribution network, which consists of a distribution server and network equipment. To date, Internet QoS has been managed by analyzing and responding to device alarms. However, response-based management makes it difficult for providers to offer a video service that will satisfy most customers because device alarms of the server and the network manage the QoS of the network layer. Moreover, the providers measure delay time and packet loss as network-layer QoS for the customer's service level agreement (SLA). However, the measurement is not based on the customer's viewpoint of the user-perceived QoS and is performed between only a small number points. Therefore, it is important to check the QoS from the customer's viewpoint and determine the locations of distribution-QoS bottlenecks. What the customer wants is to have the overall QoS managed from the video source through the network to his/her terminal. In addition, it is important to respond promptly to a user's inquiry about quality degradation. There are a lot of management objects because we assume a large-scale network. Therefore, QoS management is essential in order to detect and deal with QoS-related issues.

In this paper, we describe a QoS management system from the viewpoint of the user-perceived QoS and distribution-QoS bottleneck localization. It detects degradation in user-perceived QoS by active measurement and locates the cause and it evaluates the QoS perceived by each user by passive measurement. This system reduces operations by the maintenance staff by regularly measuring the QoS and localizing degraded-QoS sections and equipment in the network. Moreover, the providers manage the QoS and can contribute to efforts to improve user satisfaction.

† NTT Network Service Systems Laboratories
Musashino-shi, 180-8585 Japan
E-mail: tsuno.akihiko@lab.ntt.co.jp

2. QoS management requirements for video distribution services

Since services provided over the Internet generally traverse equipment belonging to multiple providers, the QoS of each participating provider (e.g., the QoS of their servers and networks) must be determined and the user must be told what is the responsibility of one provider and what is the responsibility of others. This involves localizing and determining the causes of problems affecting QoS. If the problem resides in a provider's own equipment, the provider must be able to identify the bottleneck and overcome it. The required QoS management must be able to detect the deterioration of transmission QoS over IP networks, pinpoint where in the network and/or the iDC (Internet data center) the QoS degradation is occurring, and pinpoint the network equipment causing the problem if the QoS problem lies in a network and/or iDC under the provider's own management.

i) Detecting the deterioration of distribution QoS over IP networks

When the QoS of an IP network is measured, the standard of deterioration varies depending on the type of content and type of distribution application, so it is necessary to provide different kinds of probe packets that simulate each type of content and each type of distribution application. For example, packet length, transmission interval, transmission time, and other parameters can be set as generation conditions. Moreover, it is necessary to set the corresponding thresholds.

ii) Pinpointing where in the network and/or iDC the QoS degradation is occurring

Measuring equipment is deployed at nodes between networks, and bottlenecks are pinpointed by gradually narrowing down the search range by breaking up the networks into ever smaller sections. The criteria are based on the dominant factors affecting the users' perceived QoS. Gathering QoS-related data from all equipment across the whole network is impractical as a means of pinpointing the bottleneck once degradation has been detected, simply because there is too much of it.

iii) Pinpointing the equipment causing the problem if the QoS problem lies in a network and/or iDC under the provider's own management

The QoS problem is caused by a shortage of resources. The equipment causing the problem may include various types. The evaluation criteria could be, for example, the packet loss rate, line activity, CPU (central processing unit) activity, and memory capacity. The QoS-related data is collected from the equipment in a network and/or iDC. However, a provider cannot collect QoS-related data from networks belonging to other providers.

3. Targeted network configuration and QoS measurement

3.1 Content distribution network configuration

A content distribution network consists of an iDC, the provider's own network, and interconnected networks managed by other companies (**Fig. 1**). Therefore, an end-user's PC is connected to an iDC via a sequence of networks; one or more of them are managed by the provider, but the others are operated by other companies. The iDC includes one or more content servers, a load balancer, a firewall, and various other pieces of equipment. While all networks use many routers, a provider can only obtain QoS data from the routers on networks that it controls itself. This QoS data is collected in a management information base (MIB), which can calculate packet loss ratios and line activity ratios.

3.2 QoS measurement

3.2.1 Distribution QoS and user-perceived QoS

To efficiently and accurately grasp the users' perceived QoS, it is necessary to clarify the relationship between the perceived QoS and the distribution QoS and identify the main factors causing QoS to decline.

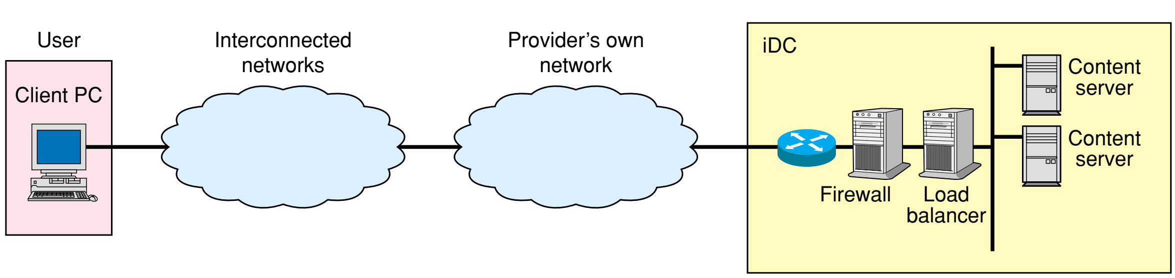

Fig. 1. Content distribution network.

Distribution QoS can be broadly divided into server QoS and network QoS. Since the primary factors contributing to reduced server QoS are CPU overload and insufficient memory, we use CPU activity ratio and remaining memory capacity as QoS criteria. And considering that the primary causes of reduced IP network QoS are router or link overload, we use the IP packet loss ratio as the QoS criterion in this case.

3.2.2 Measurement of IP network QoS and server QoS

When the QoS of a server is measured, logs that exist in the server are collected, and deterioration is detected. The relationship between the log and the image quality of each application must be examined because the log is different in each application and the threshold must be set.

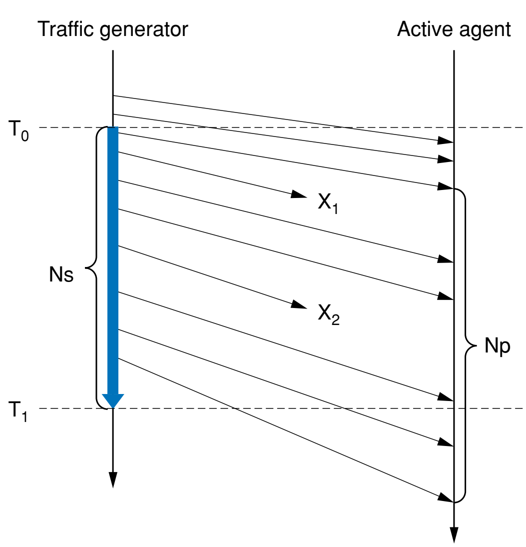

Fig. 2. Definition of packet loss ratio.

3.2.3 Definition of packet loss ratio

The packet loss ratio can be measured by an active agent, as shown in **Fig. 2**. We use *Pr* (defined by Eq. (1)) as a measure of the packet loss ratio. It is necessary to synchronize the time between the traffic generator and the active agent. Note however that if the last packet fails to reach the active agent, this impedes our ability to perform periodic measurements. We deal with this potential problem by assigning an appropriate time-out value to the last packet. If the first packet fails to reach the active agent, we can recognize the loss of the packet by counting the number of packets that reach it in the measurement time (T1 – T0 + time-out value).

$$Pr = (Ns - Np)/Ns, \qquad (1)$$

where Ns is the number of packets sent by the traffic generator between T0 and T1, Np is the number of packets received by the active agent that were sent by the traffic generator between T0 and T1, and Xi is the number of lost packets.

4. QoS management system

In a large-scale network, it is necessary to set up a lot of active/passive agents. This large number of agents enables us to localize the degraded-QoS section efficiently. Considering the effort involved, it is not realistic for us to manually set the agents, synchronize them, operate them, and collect the results. In addition, manual control would include the possibility of human error. Therefore, it is preferable to perform the setting and operation automatically from a remote site.

Our QoS management system is schematically shown in **Fig. 3**. QoS degradation is detected by peri-

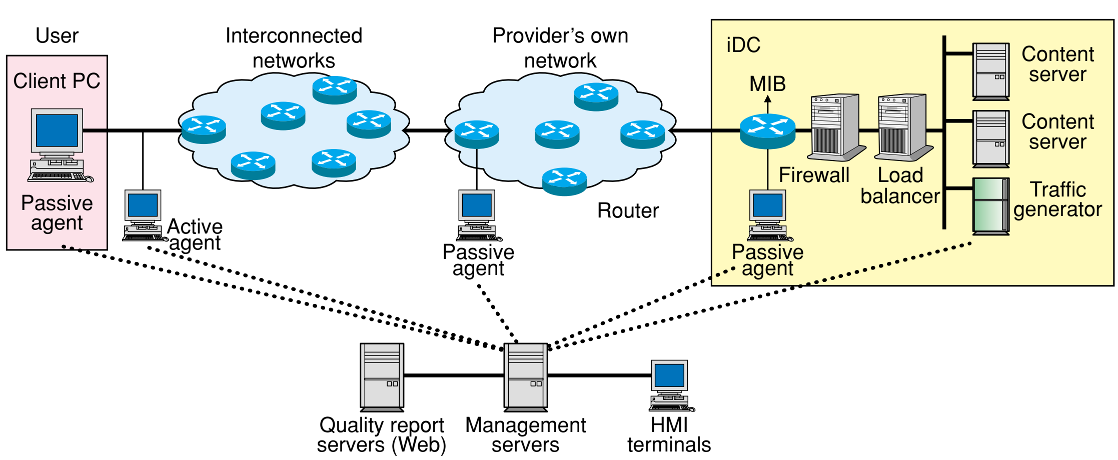

Fig. 3. QoS management system for video distribution services.

Table 1. Quality reports.

	Purpose		
	Users	Providers	iDCs
i) Specifying the distribution route	· Confirmation of service conditions · Selection of video content delivery service	· Business for users · Network management	
ii) Localizing the degraded-QoS section		· Network management	
iii) Localizing the degraded-QoS equipment		· Network management · Upgrading of equipment	

odically measuring the QoS management criteria near the users. The system consists of active agents, passive agents, traffic generators, HMI (human-machine interface) terminals, management servers, and quality report servers. They are implemented in software.

To measure the QoS, the traffic generator sends probe packets simulating actual content traffic over the network. These are detected by the active agent, which measures the packet loss ratio. The QoS of the network as a whole is inferred from the data collected by the active agent and is used in the provisioning of network resources. The passive agents are deployed in end-user PCs, in network equipment, and in the iDC and they support the customer service staff by judging the QoS perceived by the user. Based on various sources of information, user quality reports are generated by quality report servers and then provided to users, providers, and iDCs (**Table 1**).

5. Localizing distribution-QoS bottlenecks

In the provisioning of network resources, it is important to detect the occurrence of QoS degradation and localize the bottleneck as quickly as possible. Our system can estimate the QoS of the network as a whole between the traffic generator and active agents by using active measurement to detect degradation of transmission QoS. If degradation is detected, the bottleneck is localized through the following three steps:

- Specifying the distribution route
- Localizing the degraded-QoS section
- Localizing the degraded-QoS equipment

After the bottleneck has been localized, the provider can allocate additional bandwidth and resources, upgrade the equipment, or take some other appropriate action.

5.1 Specifying the distribution route

When QoS degradation occurs, it is essential to identify the route over which the content packets traveled. But in dynamic routing, cache and mirror servers are used, so the content delivery route is not constant. When QoS degradation is detected, it is necessary to specify the distribution server. Then the degraded-QoS distribution route from the distribution server to the active agent is specified by using routing information and traceroute information.

5.2 Localizing the degraded-QoS section

Next, we determine whether the degraded-QoS section is in the iDC or in the network. If the primary cause is the performance of the content server, this can be readily determined simply by scrutinizing the server log, so the first step is to analyze the server log. But if the source is in the network, the localization granularity varies depending on the network's configuration. For example, a provider can perform much finer-grained localization and narrow down to much smaller sections on networks under its own control than on networks controlled by other providers. Localization of sections is done by active measurement using packet losses, throughput, and other degradation criteria (**Fig. 4**).

5.3 Localizing the degraded-QoS equipment

The next step is to localize the equipment causing the QoS degradation in cases where the equipment is on the provider's own network. MIB data is collected from equipment (routers and other equipment having MIBs) on sections localized in the previous step, and bottlenecks are pinpointed using the packet loss ratio, line activity ratio, and other degradation criteria (**Fig. 5**).

6. QoS perceived by each user

When providers receive inquiries or complaints

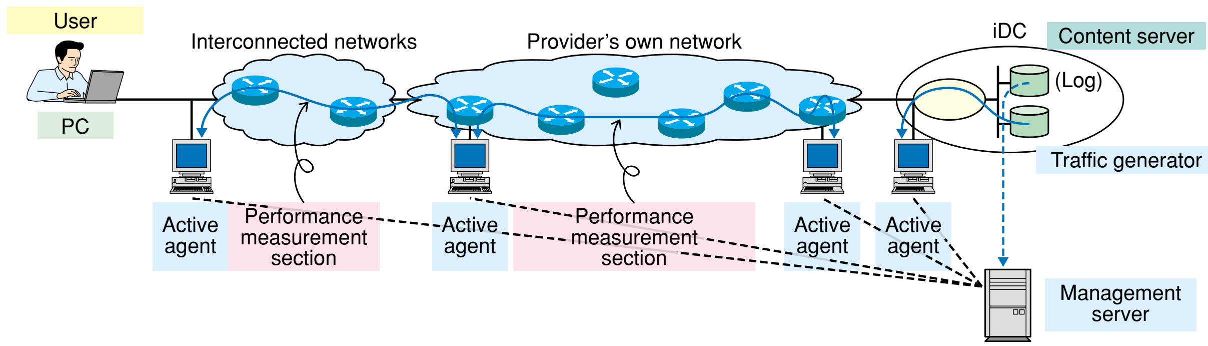

Fig. 4. Scheme for localizing degraded sections.

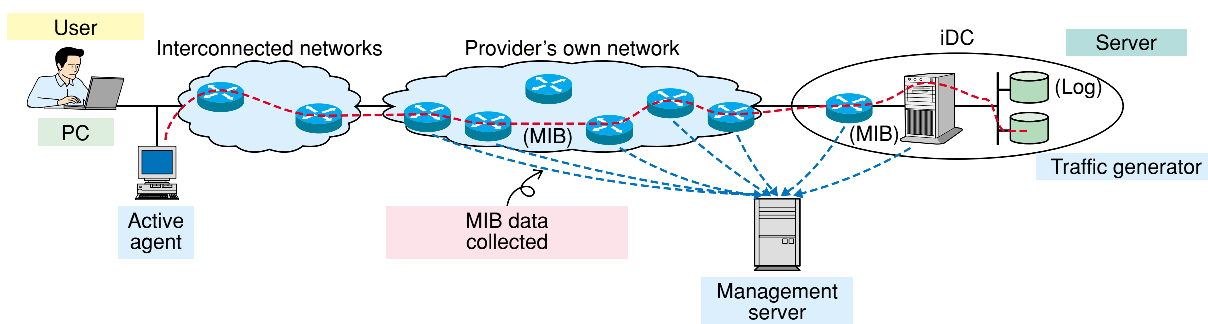

Fig. 5. Scheme for pinpointing bottlenecks.

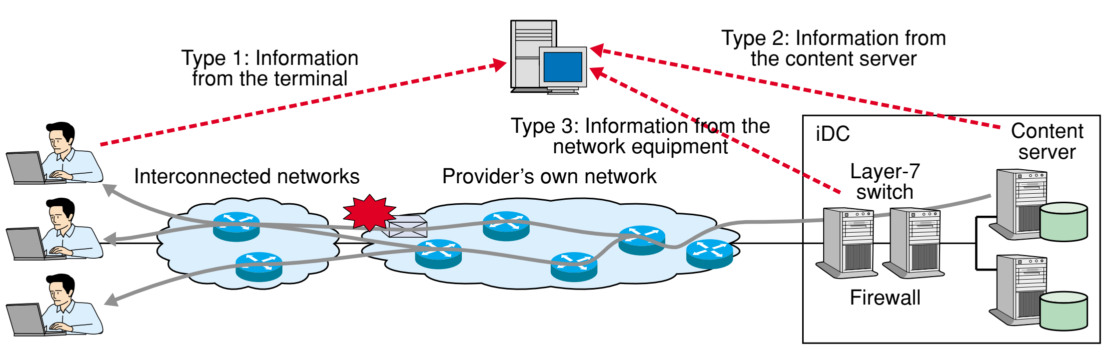

Fig. 6. Possible deployments of passive agents.

about the picture and sound quality of content, they must verify the claims, identify the cause of the problem, and respond to the customer. In some cases, the cause will be the user's terminal, but in other cases it will be the network.

To grasp the distribution QoS delivered to each user, it is not enough to know the network QoS. If packet retransmission and other control capabilities are supported in the application layer, then detailed QoS information can be obtained by passive agents. There are three possible ways of deploying passive agents, as shown in **Fig. 6**. Here, we briefly consider the features of each type.

- Type 1: A passive agent is installed on the user's PC. Its software collects the necessary QoS-related information. This permits the collection of information in the application layer, but if the information and the acquisition method depend on the distribution application, then different agent software may be required for each distribution application.
- Type 2: Passive agents are installed in the content server in the iDC. Quality information is collected using the log information stored in the content

distribution server (access and application logs). This permits state management of all users accessing the server, but has a similar problem to Type 1: if the application log is used, the log information could differ with every distribution application.

- Type 3: Passive agents are installed in network equipment such as layer-7 switches. An interface is specified between the passive agent and network equipment, and higher-layer information is collected and aggregated by an external device. This enables the status of applications to be tracked while avoiding bottlenecks.

These deployment types are compared for one agent in **Table 2**. Type 1 enables us to acquire the QoS data for each user individually. With Types 2 and 3, we must extract each user's QoS data from the QoS data of all users accommodated by the content server and by the network equipment, respectively. Therefore, we chose Type 1 considering the volume of data and the computational complexity of collected QoS data for our system.

Table 2. Comparison of passive agents.

	Volume of data (by one agent)	Data acquisition cycle (minimum)	Computational complexity
Type 1	Small	A few seconds	Small
Type 2	Large	Every 8 kilobytes	Large
Type 3	Large	A few seconds	Large

7. Comparison of QoS evaluation by active and passive measurement

In this section, we show an example that compares the use of active measurement with passive measurement for evaluating QoS criteria. In the Type 1 passive measurement described in the previous section, application-layer QoS data is obtained from the various users' PCs, and the QoS criterion can be represented by

$$Quality = (r + rr)/(r + rr + PacketLost) \times 100, \quad (2)$$

where r is the number of received packets, rr is the number of received retransmitted packets, and *PacketLost* is the number of packets not restored either by error correction in the application layer or by retransmission in the network layer.

Figure 7 shows measurement results when an active agent was deployed near the user and a passive agent was installed on the user's PC. *Pr* (defined by Eq. (1)) is the packet loss rate in the IP layer. Event 1 represents the situation where packets were continually being lost, and the *Quality* value was declining due to distribution application buffering and timeouts. Good agreement was obtained between the values measured by the active and passive agents. On the other hand, Event 2 is the situation where packet loss occurred in the network layer, but the *Quality* in the application layer fell off precipitously. Even though packet loss occurred, the *Quality* value did not decline as much as in Event 1 because the packets

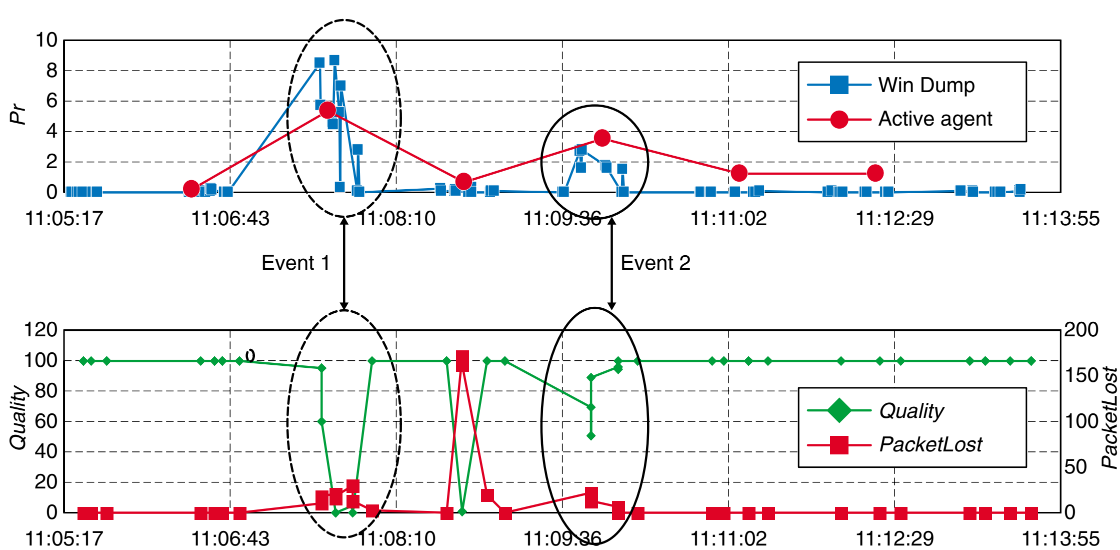

Fig. 7. Quality criteria (*Pr*, *Quality*, and *PacketLost*).

were restored by the retransmission capability of the distribution application.

8. Conclusion

After reviewing the requirements for QoS management for video distribution services, we described our QoS management system, which detects degradation in user-perceived QoS by active measurement and locates its cause and evaluates the QoS perceived by each user by passive measurement. We assumed that the system's management objects were 200 active agents, 200 distribution servers, and 1000 routers.

We described how to detect degradation by active measurement and locate its cause. Moreover, we tested a system as an example of QoS evaluation by active and passive measurement. If the application supports retransmission and other control capabilities, the combination of active and passive measurement in the application layer provides a more accurate grasp of the distribution QoS and the users' perceived QoS.

The method of detecting QoS degradation should consider the influence of active measurement on the network. In addition, it is important to understand the characteristics of the distribution application and to decide how to arrange the measurement points and set the measurement period. Moreover, it is important to use information obtained from the network element management system to collect network composition and status information effectively. We will investigate these points in future.

References

[1] T. Sakurai, S. Majima, M. Yokoyama, and H. Tokunaga, "Study of QoS management architecture in the multimedia services era," Technical Proceedings of APNOMS 2001, Sydney, Australia, Sep. 2001.

[2] M. Brodie, I. Rish, and S. Ma, "Optimizing Probe Selection for Fault Localization," 12th IFIP International Workshop on Distributed Systems: Operations & Management DSOM2002, Nancy, France, October 15-17, 2001.

[3] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP Throughput: A Simple Model and its Empirical Validation," Proceedings of ACM SIGCOMM98, Sep. 1998.

Akihiko Tsuno

Engineer, Network Software Service Project, NTT Network Service Systems Laboratories.

He received the B.E. and M.E. degrees in information computer engineering from Niigata University, Niigata in 1999 and 2001, respectively. He joined NTT in 2001. Since then, he has been engaged in research on quality management systems. He is currently engaged in the development of session control software.

Yoshihiro Otsuka

Senior Research Engineer, Supervisor, Network Software Service Project, NTT Network Service Systems Laboratories.

He received the B.E. degree in electrical and electronic engineering from Kanazawa University, Ishikawa in 1983 and the M.E. degree in physical electronics from Tokyo Institute of Technology, Tokyo in 1985. Since joining NTT in 1985, he has mainly been engaged in research on broadband switching systems and network management systems. He is currently working on the modeling of network management functions and multi-layered network management systems. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan.

Souhei Majima

Senior Research Engineer, Supervisor, Network Software Service Project, NTT Network Service Systems Laboratories.

He received the B.E. and M.E. degrees in information science from Fukui University, Fukui in 1982 and 1984, respectively. He joined NTT Laboratories in 1984. He has been involved in the development of a computer-aided software development environment (CASE) system and a switching equipment management system for large-scale networks. He is currently engaged in research on an NGN operation support system. He is a member of IEICE and the Information Processing Society of Japan.