Unavailability Evaluation Method for Communication Network Management

Tatsuya Matsukawa, Nippon Telegraph and Telephone Corporation

Hiroyuki Funakoshi, PhD, Nippon Telegraph and Telephone Corporation

Key Words: unavailability evaluation, communication network, reliability management

SUMMARY & CONCLUSIONS

In this paper, a method of unavailability evaluation for communication network management is proposed. This method supports the management of large-scale communication networks provided to end users. The unavailability model for communication networks and the evaluation procedure is described. The time trend of network reliability can be understood by considering the characteristics of unavailability over short periods. Examples of evaluation of the network unavailability and management scenarios in a communication network are presented. Our reliability modeling and evaluation method can be flexibly applied to various types of network architectures and has been used successfully in our work.

1 INTRODUCTION

As broadband capability in internet protocol (IP) networks has developed, IP telephone, IP-TV, and other multimedia services have been provided to end users. The Internet has become an important infrastructure for social and economic activity. Therefore, reliability and maintainability (RAM) has become a critical requirement for communication networks and telecommunication carriers have to provide communication networks with high service availability.

Most end users use services from their Internet service provider (ISP) and application service provider (ASP) through the carrier network. Telecommunication carriers need a framework to maintain a level of network reliability to meet their requirements. Service level agreement (SLA) and operation level agreement (OLA) are approaches for managing network reliability [1]. In Refs. [2] and [3], the procedure to set objective availability is studied. In Ref. [4], the risk for SLA design is discussed. In these studies, availability is used as a common measure for telecommunication networks. It is also defined as an IP packet performance measure [5].

Evaluation methods for network reliability have been studied. For example, the Network Reliability Council in the USA assesses the service reliability provided by a telecommunication carrier over a certain period by using an outage impact measure (OIM), which quantifies the effect of service outages [6]. NTT, a telecommunication carrier in Japan, has a reliability standard that specifies the reliability targets for the public switched telephone network (PSTN). Fairly clear reliability targets for PSTN are in Ref. [7]. Moreover, a tool for network reliability design has been developed [8][9].

However, there are several problems in applying these results to reliability management of large-scale communication networks such as IP networks.

- IP networks are constructed of many types of network elements, and network structures cannot always include redundancies because of cost constraints and technical limitations. For example, an interface in a router is sometimes only a single point (Fig. 1), and a component failure may cause the service outage. As the reliability of all components cannot be estimated exactly before the operation, the expected levels of network reliability cannot be achieved in the network-design phase. These levels also have to be monitored and maintained in the network-operation phase [10].
- In carrier-grade networks, the number of users increases sharply. Network equipment is added and the network structure is sometimes changed after the service has been deployed. The failure characteristics of network elements change over a short time, and the characteristics of service outage time also may change. Therefore, the reliability model assumed in the network-design phase is not always applicable to evaluation in the network-operation phase.

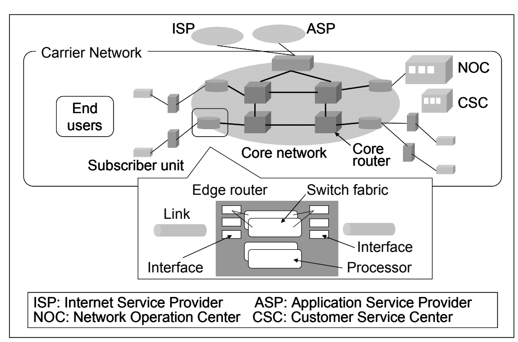

Fig. 1 Example of communication network and elements

Previous studies aimed at maintaining the level of reliability in the network-design phase and are not suitable for applying to the management of network reliability. These studies modeled only the specific network elements, and the

1-4244-2509-9/09/$20.00 ©2009 IEEE

characteristics of network reliability were not investigated. Therefore, they are not suitable for applying to the management of overall networks whose reliability and conditions are changed.

To solve these problems, we developed an unavailability evaluation method for communication networks and investigated unavailability in a large-scale communication network that is provided for end users. This paper describes a model for network unavailability considering short-time characteristics, an evaluation procedure for the management cycle, and evaluation examples using actual failure data.

2 UNAVAILABILITY EVALUATION BY ANALYZING FAILURE DATA

Network reliability management is a cycle of reliability activities. After collection of the failure data for a certain period, analyzing and evaluating the reliability measures are the key process in the cycle. Corrective actions are performed in accordance with the results, and the level of network reliability is then monitored to measure the effectiveness of the actions. The network reliability design is a part of the reliability management framework.

The following information about failures is usually collected and stored in the database systems for network operation.

- Name and type of failed network element
- Date and time of failure
- Date and time of recovery
- Affected services or number of users affected by the failure

The failure data from communication networks is collected and aggregated at the network operations center (NOC). At the customer service center (CSC), data such as the number of users is collected. If the unavailability in a targeted service can be evaluated from these data, the level of network reliability can be managed. The following are requirements regarding the management cycle.

- Network availability and unavailability are evaluated by analyzing the updated failure data, which involves all unplanned service outages. It is simple to calculate the network unavailability from the failure data because unavailability is calculated as 1 minus availability.
- Network unavailability for the entire network provided by telecommunication carriers is evaluated, as the network unavailability shows the service level for users.
- The time trend of network unavailability is sequentially estimated so that the network operators can use the results to quickly review the actions for reliability improvement.

3 UNAVAILABILITY EVALUATION METHOD

In this section, the network unavailability model and evaluation method are shown. To evaluate the network unavailability, the model for the time trend of unavailability is described. Moreover, a criterion considering the unavailability distribution is proposed.

3.1 Network unavailability model

In large-scale communication networks, the effects of failures greatly differ as the number of users affected by a failure depends on the failed network element. One of the measures for describing reliability in these networks is unavailability for an arbitrary user, which is defined by the time ratio of unavailable time for a user to observed time. The unavailable time for a user is calculated as the expected value of outage time for the network element. Moreover, as unavailability due to a failure in a network element for a user is expressed as the time ratio of outage time per user to observed time, the unavailability due to the ith failure is given by d_i/r_iT.

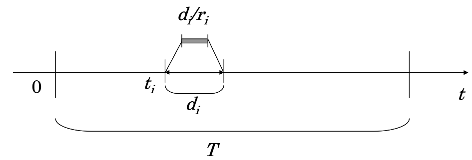

Fig. 2 Definitions of parameters

Network unavailability is approximated by the sum of the value of unavailability in each network element that is not redundant. Therefore, if each failure occurs independently, network unavailability is expressed as the sum of the unavailability due to each failure. The unavailability in a given time interval T is represented by the following equation.

$$u = \frac{1}{T} \sum_{i \in [t_i \le T]} \frac{d_i}{r_i} \quad i = 1, 2, 3, \cdots \tag{1}$$

The definitions of parameters are illustrated in Fig. 2 and described in Table 1. If accurate information about the total number of ith failed elements r_i is not obtained, r_i is replaced by N_i/n_i, which corresponds to the total number of ith failed elements.

Parameter	Definition
i	failure number
d_i	outage time of ith failure
r_i	total number of ith failed elements (total number of type of elements that failed equipped in network)
t_i	Time point when ith failure occurs
n_i	number of users affected by ith failure
N_i	total number of users in targeted service at time point of ith failure
j	number of calculations
c_j	jth time point when calculation is performed (c_0=0)
u_j	jth calculated value of unavailability
k, l	number of management period
T_k	end point of time for kth management period
$\bar{u}$	value of evaluation criterion

Table 1 Important parameters and definitions

3.2 Network unavailability evaluation

If T in eq. (1) is small, network unavailability for a short time interval is described by the jth point of time c_j and the jth calculated value of unavailability u_j. The time trend of network unavailability is described by the following coordinates:

$$\left[c_j, \frac{1}{c_j - c_{j-1}} \left(\sum_{i \in [c_{j-1} < t_i < c_j]} \frac{d_i}{r_i} \right) \right] \quad j = 1,2,3\cdots \tag{2}$$

By use of these coordinates, the time trend of unavailability is evaluated.

The calculated value of unavailability is evaluated by comparison with the criterion value. For example, if the calculated value of unavailability u_j meets the following inequality, network availability is low at the time point of the jth calculation.

$$u_j > \bar{u} \qquad j = 1, 2, 3, \cdots \tag{3}$$

To manage the network reliability, the management periods are set in the whole measured period. The management period is set in accordance with the cycle of the network operation, and the span is a few months or a quarter. In this period, some values of unavailability are calculated and the set of these values is compared to the value of the criterion. For example, the set of values of calculated unavailability in the kth period is compared with the value of the criterion. If the jth calculated values of unavailability meets the following inequality, network availability is low at the time point of the jth calculation in the kth period.

$$u_j > \bar{u} \qquad j \in [T_{k-1} < c_j \le T_k] \tag{4}$$

Here, T_k are the end points of time for the kth period.

As a procedure to characterize the time variation of unavailability, the frequency at which the calculated value of unavailability exceeds the criterion value in the period is counted. This is counted in each period and is compared with that in other periods. As an example, results in the lth period and the kth period are compared using the following inequality.

$$\sum_{j \in [T_{l-1} < c_j < T_l]} 1_A(j) > \sum_{j \in [T_{k-1} < c_j < T_k]} 1_A(j) \tag{5}$$

Here, $1_A(j)$ is the indicator function of a subset A, which meets the following equality.

$$1_A(j) = \begin{cases} 1 \text{ if } j \in A \\ 0 \text{ if } j \notin A \end{cases} \tag{6}$$

A subset A is defined by $A=\{j|\ u_j > \bar{u}\}$. T_l and T_k are the end points of time for the lth period and the kth period, respectively, in Fig. 3. The right side in eq. (5) shows the frequency that unavailability exceeds the criterion value. Therefore, inequality in eq. (5) means that the network in the kth period is more stable than that in the lth period.

Thus, the time trend of unavailability is evaluated by the following procedure.

Step 1 Set the time interval of calculation.
Step 2 Calculate the network unavailability from eq. (2).
Step 3 Set the criterion value for evaluation.
Step 4 Count the frequency for some periods from eq. (6).
Step 5 Compare the results of step 4.

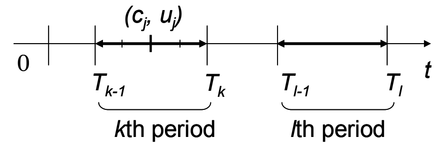

Fig. 3 Illustration of parameters

Moreover, in the case of evaluating the unavailability cyclically, the evaluation criterion is set sequentially by using the unavailability in the previous period. The evaluation criterion for the kth period is represented as a function of unavailability in the k-1th period as follows.

$$\bar{u} = h(u_j) \quad j \in [T_{k-2} < c_j \le T_{k-1}] \tag{7}$$

3.3 Evaluation criteria

As a general function for the evaluation criterion, the distribution of unavailability in the previous period is used. As the unavailability is defined by the ratio of the unavailable time to the measured time, the distribution of unavailability is calculated by dividing the whole period into short periods and calculating the ratio for each short period. From the following characteristics of network reliability and network structure, the model for the probability distribution of unavailability is selected.

- Outage times are log-normally distributed.
- Time between failures is exponentially distributed.
- Number of equipment for each type has logarithmic distribution.

From the above assumptions, the ratio of the outage time to the number of elements d_i/r_i has the characteristics of log-normal distribution. Therefore, a distribution of unavailability is estimated by assuming this distribution. The cumulative distribution of unavailability is given by the following equation.

$$F(u_j) = \Phi\left(\frac{ln(u_j) + \mu}{\sigma} \right) \tag{8}$$

Φ is a cumulative distribution function of a normal distribution, and μ and σ are average and standard deviations of a normal distribution, respectively. From eq. (8), the percentile function of a log-normal distribution is given by the following equation.

$$G(p) = exp\left(\sigma \Phi^{-1}(p) + \mu\right) \quad 0 \le p < 1 \tag{9}$$

Here, Φ^{-1} is a percentile function of a normal distribution. Therefore, the criterion value for evaluating the unavailability $\bar{u}$ is set by the following procedure.

Step 1 Calculate μ and σ for u_j.
Step 2 Set probability p.
Step 3 Calculate $\bar{u} = G(p)$ from eq. (9).

The given value p is set in this procedure. When the variation of unavailability is large, p is also set as a large value.

4 EXAMPLES OF NETWORK UNAVAILABILITY EVALUATION

In this chapter, examples of our method are shown. These results were obtained from failure data that was collected over a year in an IP network. The network consisted of more than ten types of network elements, and the number of failures was a few thousand. The failure data included only unplanned service outages and information about failure date and time, recovery date and time, and the type of network elements. The failure data was analyzed by supplementing the data of the number of affected users, the total number of users of the service, and the total number of elements of each type.

4.1 Example of unavailability analysis

The result of an unavailability analysis obtained from eq. (2) is shown in Fig. 4. The time intervals for calculation were 3 (s=3) and 10 (s=10) days. In these results, unavailability in the 3rd and 4th quarters was smaller than that in the 1st and 2nd quarters. For 3-day time intervals, the difference in maximum and minimum unavailability was greater than that for 10-day time intervals. At 10-day time intervals, unavailability in the 3rd and 4th quarters stabilized.

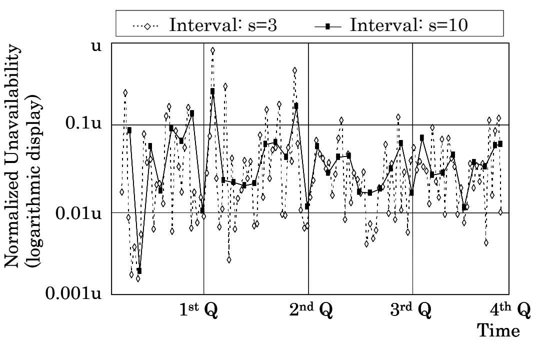

Fig. 4 Example of unavailability analysis

4.2 Example of unavailability evaluation

The same data as in section 4.1 was used to show an example of an unavailability evaluation (Fig. 5). The time interval for calculation was 7 days (s=7). The percentile of unavailability was estimated from eq. (9) by using a year of failure data. As evaluation criteria $\overline{u}$, 75th percentile (p=75) and 90th percentile (p=90) of calculated values of unavailability were set. In the 1st and 2nd quarters, the unavailability varied considerably with time, the unavailability exceeded the baseline at some points of time, and 6 calculated values of unavailability exceeded the baseline (p=90). On the other hand, in the 3rd and 4th quarters, calculated values of unavailability did not exceed the baseline (p=90).

Another example of unavailability evaluation is shown in Fig. 6. The time interval for calculation was 7 days (s=7). The percentile of unavailability was estimated from the failure data in the 1st and 2nd quarters. A 75th percentile (p=75) and 90th percentile (p=90) were set as evaluation criteria. In the 1st and 2nd quarters, 8 calculated values of unavailability exceeded the baseline (p=90). On the other hand, in the 3rd and 4th quarters, 1 calculated value of unavailability exceeded the baseline (p=90).

These results demonstrate that network unavailability decreases and stabilizes with time. By our method, the characteristics of unavailability for a given time can be evaluated and the time trend can be quantified.

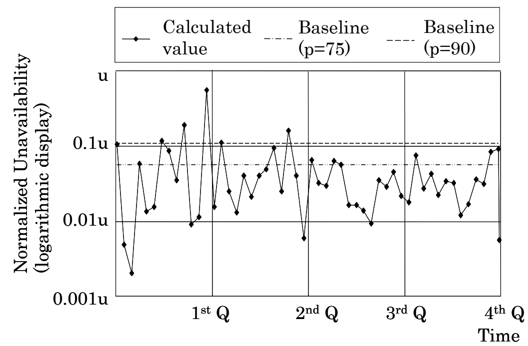

Fig. 5 Example of unavailability evaluation

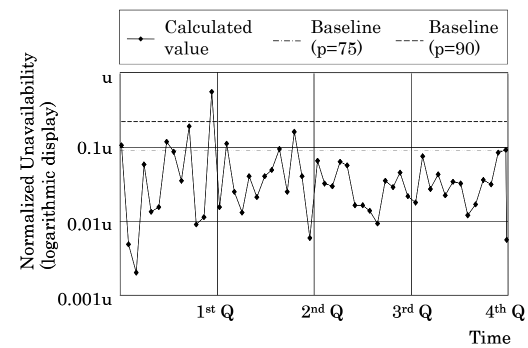

Fig. 6 Example of unavailability evaluation

4.3 Verification of characteristics of unavailability distribution

The estimation results of empirical distribution and log-normal distribution for the unavailability are shown in Fig. 7. The data collected for a year was divided into 7-day intervals, and the distribution of the unavailability was calculated from that data. Moreover, to verify the characteristics of the distribution of unavailability, a normal probability plot of the logarithm of unavailability was made (Fig. 8). The points on this plot form a nearly linear pattern, which indicates that the log-normal distribution was a good model for this data set.

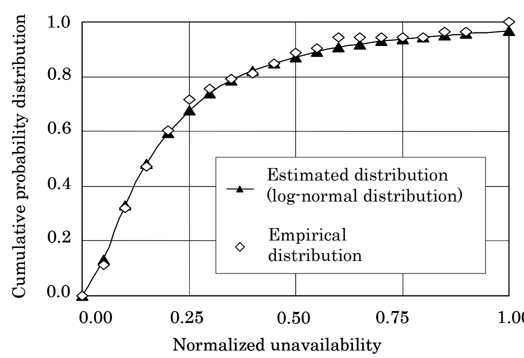

Fig. 7 Estimation of unavailability distribution

As additional verification, a chi-square test was applied to this data set. In this case, the test had two degrees of freedom and a chi-square value of 0.28. These conditions are statistically significant with a 0.01% confidence level. The results demonstrate that the log-normal distribution is a good model for setting the criterion and the characteristic of unavailability can be simply evaluated by using the information on that distribution.

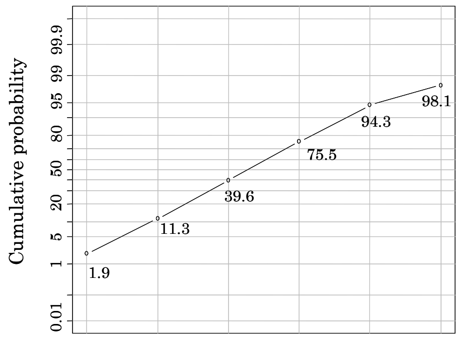

Fig. 8 Normal probability plot of logarithm of unavailability

5 APPLICATIONS TO NETWORK MANAGEMENT

When the results for some periods are compared, the level of the network unavailability can be evaluated as shown in section 4.2. In this section, network management scenarios are shown.

5.1 Analysis of long-term characteristics

Another description of unavailability for the same data as that in section 4.1 is shown in Fig. 9. The unavailability for a given term was evaluated from eq. (1). The lower graph in Fig. 9 is the trajectory of the date of the ith failure and the cumulative value of unavailable time per user (t_i, $\sum_{1\leq m\leq i} d_m / r_m$). The gradient of the curve in the lower graph equals the unavailability. The upper graph in Fig. 9 shows the ith failure date and the relative value of unavailability for the ith failure (t_i, d_i/r_i) as a scatter plot. This visualizes the effect of one failure on the network unavailability. In this example, the unavailability in the 2nd quarter is almost twice that in the 1st quarter. Failures in the 2nd quarter cause the rapid increase of network unavailability.

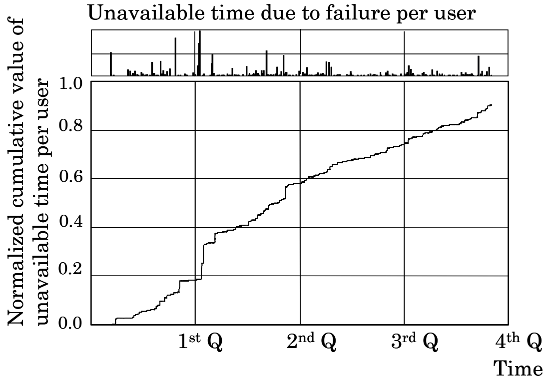

Fig. 9 Analysis of long-term characteristics

5.2 Unavailability of specific types of network elements

To review the management policy for network elements, we investigated the trend of unavailability for specific types of elements. An analysis of three types of network equipment in this network is shown in Fig. 10. These results were obtained by the same approach as that in section 5.1. The three curves in Fig. 10 show the cumulative values of the unavailable time per user. In Table 2, the results of linear regressions for the curves are shown. The unavailability of type-3 network equipment is highest and stays at constant values. In this example, though the unavailability of type-2 decreased, failures in the 1st and 2nd quarters have to be investigated. Reviewing the recovery procedures and shortening the outage time should maintain the level of network availability .

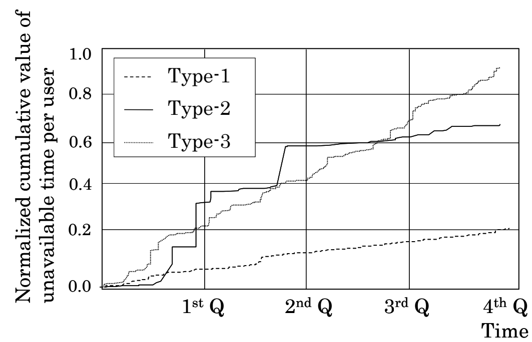

Fig. 10 Unavailability for three types of network elements

Type	Linear coefficient (normalized value)	Coefficient of determination
1	u	0.97
2	3.2u	0.84
3	3.9u	0.99

Table 2 Results of linear regression

5.3 Unavailability characteristics based on number of affected users

To investigate the unavailability characteristics for the number of users affected by a failure, the failure data is grouped into three classes based on the number of users affected by a failure. The classes are “small”, “medium”, and “large”. We performed the same analysis as that in section 5.2 and obtained the three curves in Fig. 11. Each class included different types of network elements. Some types of network elements were included in two classes, and a type did not always correspond one-to-one with a class because a greatly differing number of users can be affected by failures in the same types of elements.

The unavailability of the small class is low. The unavailability of the medium class decreased while the unavailability of the large class stayed high. This is because the total number of users in the service increased during the measured period, so the number of users affected by a failure increased. In this case, the level of network availability is expected to be improved by adding equipment and reducing the number of users affected by a failure.

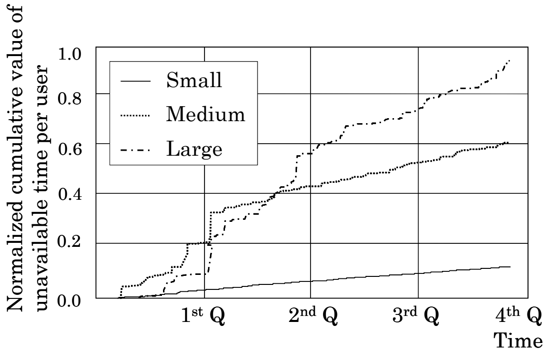

Fig. 11 Unavailability for classes of failure

Class	Linear coefficient (normalized value)	Coefficient of determination
Small	u	0.99
Medium	4.4u	0.89
Large	7.8u	0.95

Table 3 Results of linear regression

6 RESULTS

We have proposed a method for evaluating the unavailability of communication networks. This method helps network operators to analyze the time trend of unavailability, particularly in large-scale communication networks. We developed a software tool for unavailability evaluation by implementing the above method among others. This tool has been developed in Japanese by using statistical software. By inputting the failure data to this tool, the network unavailability is automatically calculated and graphically displayed.

As this unavailability model and evaluation procedure does not depend on the network architectures, this method can be applied to various types of communication networks. Moreover, this method is especially effective for management of large-scale communication networks because a large amount of failure data can be collected from their networks.

REFERENCES

1. http://www.nextslm.org/
2. L. Zhou and W. Grover, “A Theory for Setting the "Safety Margin" on Availability Guarantees in an SLA,” in Proc. DRCN2005, pp. 403–409, Oct. 2005.
3. R. Clemente, M. Bartoli, M. Bossi, G. D’orazio, and G. Cosmo, “Risk Management in Availability SLA,” in Proc. DRCN2005, pp. 403–409, Oct. 2005.
4. C. Kelvin/K. Chu, H. Pant, S. H. Richman, and P. Wu "Enterprise VoIP Reliability," in Proc. IEEE NETWORKS, pp. 1–6, Nov. 2006.
5. Y.1540. Internet protocol data communication service — IP packet transfer and availability performance parameters. ITU-T, 2002.
6. E. S. Tollar and J. Benett, “Network Outage Impact Measures for Telecommunications,” in Proc. IEEE Computers and Communications, pp. 120–126, June 1995.
7. S. Nojo and H. Watanabe, “Incorporating Reliability Specifications in the Design of Telecommunication Networks,” IEEE Com. Mag., Vol. 31, No. 6, pp. 40–43, June 1993.
8. T. Abe, M. Hayashi, S. Nojo, “A Software Tool to Support the Reliability Design and Evaluation of Telecommunication Networks” IEEE JSAC., Vol. 12, No. 2, pp. 345–354, Feb. 1994.
9. H. Watanabe, M. Kudo, K. Kawanishi, and K. Yamasaki “A Reliability Design Method for Private Networks,” in Proc. Reliability and Maintainability Symposium, pp. 237–243, 2003.
10. C. R. Johnson, Y. Kogan, Y. Levy, F. Saheban, and P. Tarapore, “VoIP Reliability: A Service Provider’s Perspective,” IEEE Com. Mag., Vol. 42, Issue 7, pp. 48–54, July 2004.

BIOGRAPHIES

Tatsuya Matsukawa
Nippon Telegraph and Telephone Corporation (NTT)
9-11, Midori-Cho 3-chome

matsukawa.tatsuya@lab.ntt.co.jp

Tatsuya Matsukawa received B.S. and M.S. degrees from Osaka University in 2000 and 2002 respectively. He joined NTT in 2002 and has been engaged in the research and development of reliability evaluation and management methods for IP networks. He is currently a research engineer of NTT Service Integration Laboratories. He received the young researchers’ award and the best paper award of IEICE in 2007 and 2008.

Hiroyuki Funakoshi
Nippon Telegraph and Telephone Corporation (NTT)
9-11, Midori-Cho 3-chome

funakoshi.hiroyuki@lab.ntt.co.jp

Hiroyuki Funakoshi received B.S. degrees in integrated arts and sciences and an M.S. degree in information engineering from Hiroshima University in 1996 and 1998 respectively. He jointed NTT in 1998 and has been engaged in research and development of reliability management for the Next Generation Network. In 2007, he received a Ph.D. degree from Waseda University. He is currently a research engineer of NTT Service Integration Laboratories. He received the best paper award of IEICE in 2008.