Chapter 32

ANALYSIS OF THE RELATION BETWEEN SERVICE PARAMETERS FOR SERVICE LEVEL MANAGEMENT AND SYSTEM UTILIZATION

Masaharu Akatsu, Shoji Konno, Norihisa Komoda

Abstract: System utilization was investigated with the aim of selecting appropriate service parameters for defining the quality of information systems such as SLA (service level agreement). This investigation is based on the hypothesis that parameters having a strong influence on user satisfaction depend on how the system is used. Accordingly, the parameters for availability and responsiveness, which are known as the two major factors in SLA, were investigated. First, it is explained that parameters having a strong impact on user satisfaction are not common among information systems. Next, three parameters for availability (mean time between halt, mean time to restore, and both of them) and ones for responsiveness (mean time, maximum time, and variance) were introduced. To analyze the relation between these service parameters and system utilization, a questionnaire about system utilization was designed and administered to several experienced system engineers. Quantification theory type II was applied to the results of this questionnaire, and the validity of the hypothesis was demonstrated. The results also classified information systems into six groups in terms of availability and responsiveness. The characteristics of each system group were also clarified.

Key words: service level management, user satisfaction, availability, responsiveness, quantification theory type II

1. INTRODUCTION

Accompanying the rise of IT service providers such as ASPs (application service providers) and iDCs (Internet data centers), it has become popular to define the quality of information systems as SLA (service level agreement) (Muller, 1999; Hiles, 1999; Lewis, 1999). SLA is based on the concept that

information systems do not only provide functions but also provide services to users. Information systems are the key factors to make or break a business's success; therefore, they are managed from the viewpoint of business managers and system users. This tide affects systems management IT professionals perform. Traditional systems management is concerned primarily with the health of the information systems and networks. On the other hand, it is becoming common to manage systems from the users' viewpoint in consideration of the impact of system quality on business, such a management style is called SLM (service level management).

The concept of "service level" is widely pervasive; however, its practice is still on the road to full realization. For example, there are no explicit criteria for the selection of service parameters. Some industry groups such as the ASP Industry Consortium and the Information Technology Association of America have presented SLA templates and guidelines for SLA. They provide a variety of service parameters, but they are little use for selecting parameters when SLA is individually defined. As a result, most service providers present almost the same SLAs. On the other hand, the true goal of SLA is to guarantee a valuable service to users. It is therefore desirable to identify service parameters that are strongly related to user satisfaction. "Cookie-cutter" parameters do not always satisfy users. The service parameters concerning each system should be different. This would lead to a way for service providers to differentiate themselves.

Over the years, numerous studies on user satisfaction have been conducted. Doll and Torkzadeh (1988) developed a standard measure of end-user satisfaction. Adelman et al. (1985) tried clarifying factors users consider most or least important when evaluating information systems. Jiang and Klein (1999) demonstrated that important factors for users change with system types. The assumption that evaluation factors can change according to system types is also applied to classification of software (Glass and Vessey, 1995).

The factors these studies have considered for system evaluation are almost the same, for example, reliability, performance, usability, accuracy. They are also considered from the viewpoint of the service level management (Strum et al., 2000; Park et al., 2001). In addition, the findings of one survey revealed that availability and performance are indispensable factors in SLA (Nextslm.org, 2000).

In spite of this result, application of these earlier works to SLA/SLM is not sufficient. Most of the research addresses user satisfaction, concentrating on evaluating important factors rather than on parameters by which the factors are measured. In the case of SLA/SLM, we need to define not only evaluation factors, but also parameters that enable us to evaluate service quality quantitatively. For example, it is not sufficient to say that availability

strongly affects on user satisfaction: we must consider user satisfaction including parameters for availability.

This paper deals with this issue on availability and responsiveness, which are two major factors in SLA. First, we explain that the parameter for each factor which strongly affects on user satisfaction is not necessarily the same. It is different for each system. So, we provide three parameters for availability and responsiveness, respectively. Next, we frame a hypothesis that parameters having a strong influence on user satisfaction depend on how the system used. To validate the hypothesis, a questionnaire on system utilization is constructed. The questionnaire was administered to several experienced system engineers. The answers are analyzed by Quantification theory type II, and the validity of the hypothesis is demonstrated. We also classify information systems into three groups to determine the parameters referred in SLA. Finally, we show how our results are used in constructing SLA/SLM.

2. SERVICE PARAMETERS FOR SERVICE LEVEL MANAGEMENT

2.1 Service Parameters for Availability

Availability is generally defined as the percentage of the time that a service is available for use. In other words, availability is calculated from Equation (1) as follows.

$$\text{availability} = \frac{\text{mean time between halt}}{\text{mean time between halt} + \text{mean time to restore}} \quad (1)$$

This equation shows that availability is determined by two parameters; mean time between halt and mean time to restore. Figure 1 expresses the relation between these two parameters in detail. Services with the same availability are located on a straight line which passes through the origin, so when availability becomes high, the inclination is large. Suppose that a service is plotted at point *a* in Figure 1. All parameters are improved as long as point *a* moves in the direction of domain A. This change is always preferable. On the contrary, moves in other directions cannot be concluded simply whether the change is desirable from the viewpoint of user satisfaction. One typical example is that the case where one of two mean

times improves but another gets worse. It depends on the service type whether this change is good or bad.

Figure 1. Relation among parameters for availability

Let us consider an example of this subtle situation. Suppose there are two service types. Service type α does not stop for 30 days, and then stops all day long on the 31st. Service type β repeats 7.5 hours work with halts of 15 minutes in between. Although the availability of both service types is equal, that is, about 97%, user satisfaction with each service is not necessarily the same. Consider a communications service via an in-house web portal, which employees use once a day. If this service belongs to type α , user satisfaction may be low because there is a day when employees cannot access the service. On the other hand, if the communications service is type β , user dissatisfaction is relatively low because employees can find a time when the service is available on any day. A movie distribution service is an opposite example. This service is continuously used for a couple of hours, so user satisfaction with type β is lower than that with type α . This is because the probability of service halts during use is higher in the case of type β .

The two above scenarios lead to the following conclusion. Availability defined by Equation (1) is not sufficient for an index for evaluation of user satisfaction. Certain parameters for calculating availability must also be considered. Accordingly, we have studied three possibilities for the parameters expressing the availability factor: mean time between halt, mean time to restore, and both of them.

2.2 Service Parameters for Responsiveness

Mean response time is a typical parameter for responsiveness. However, many studies have shown that satisfaction with responsiveness is not a simple matter of speed of delivery (Butler, 1983; Culwin and Faulkner, 2001). For example, low variance and predicted response can be a key to

satisfaction (Ritchie and Roast, 2001). Moreover, an extremely slow response has a bad influence on users (Endo and Seltzer, 2000). Hence,

Figure 2. Relation among parameters for responsiveness

lowering maximum response time contributes to the improvement in user satisfaction. Even if mean response time gets worse, the variance and the maximum response time can be improved. Figure 2 shows the situation briefly. For simplicity, the figure assumes that there are only two response times (x_1, x_2 ; $x_1 \leq x_2$) and that both of them occur with the same frequency. Services with the same variance are plotted in a straight line with the slope 1. Services with the same mean response time are plotted in a straight line with the slope -1. Suppose that a service is plotted at point b in Figure 2. Move in the direction of domain B is always preferable, but it is not necessarily appropriate to suggest that moves in other directions are good or bad. Hence, it is not simple to select a parameter to be improved, that is, the most important parameter. The importance of each parameter is independent of user personalities (Guynes, 1988), but depends on system utilization. Accordingly, we prepare three parameters for the response factor: mean time, maximum time, and variance.

3. INVESTIGATION ON SYSTEM UTILIZATION

It is desirable to select a parameter that is the most influential to user satisfaction from the parameters mentioned in the previous section. One of the simplest ways to select a parameter is to make users choose a parameter that they regard as the most important. However, this is not appropriate because users tend to think all parameters are important.

Then, we pay our attention to system utilization. This is based on a hypothesis that parameters having a strong influence on user satisfaction

depend on how the system is used. If the hypothesis is correct, investigating the utilization of many systems and analyzing them statistically reveals significant relations between system utilization and important parameters. By using such statistical analysis, the parameters that should be selected as a suitable measure for user satisfaction can be identified.

To validate the hypothesis, a questionnaire on system utilization is constructed. First, utilization time, frequency of use by the typical users, and available time are considered as items that can distinguish the services mentioned in the previous section. Next, people capable of using systems, number of users, and function the system provides are added in a questionnaire. These items are based on “reach” (the locations and people the system is capable of connecting) and “range” (functionality in terms of business activities) that Weill and Broadbent (1998) introduced with an aim to evaluate the value of infrastructure. Moreover, impact on business incurred by the loss of the system and peak load time is added. They are expected to have a strong relation with availability and responsiveness, respectively. Table 1 shows a questionnaire about system utilization. The answer to each question item is split into three or four categories in advance, and respondents select one of the prepared answers.

4. ANALYSIS BY QUANTIFICATION THEORY TYPE II

To analyze the relation between service parameters and system utilization, we gave the questionnaire survey to several experienced system engineers. They answered the questions in Table 1 for each system they had developed. They also chose parameters that they thought are the most influential on the user satisfaction in terms of availability and responsiveness. All data were given as categorical variables. Thus, Quantification theory type II can handle the analysis of the relation between important parameters and system utilization (Komazawa, 1982). In fact, the answers to the eight questions in Table 1 were taken as predictor variables, and the replies to parameters for availability and responsiveness were taken as criterion variables.

Table 1. Questionnaires on system utilization

Item		Category
1	people capable of using systems	1. designated members 2. rank-and-file employees 3. anyone
2	number of users	1. less than 10 2. from 10 to 100 3. from 100 to 1000 4. over 1000
3	available time	1. 24 hours a day, everyday 2. predetermined period of time (ex. business hours)
4	peak load type	1. peak load at a specific day or time 2. peak load at an unpredictable time 3. no peak load
5	frequency of use by the typical users	1. everyday 2. a couple of times a week 3. once a week or less
6	utilization time	1. a few minutes 2. a few minutes per access and repeating it 3. over several tens of minutes without a break
7	function the system provides	1. send messages 2. access to information 3. perform transactions
8	impact on business incurred by the loss of the system	1. direct financial loss such as decline of sales 2. indirect financial loss such as operating inefficiency 3. no financial impact

The results of the questionnaire provided data on 42 systems. There were a variety of systems, such as workflow, CRM, financial application. Quantification Theory type II was applied to 37 data sets selected at random. The five remaining data sets were used for validation. Table 2 and Table 3 list the results of this analysis and Figure 3 and Figure 4 plot 37 sample scores as scatter plots for discrimination. The discriminant ratios (percentages correctly classified) are 92% for availability and 95% for responsiveness. Table 4 lists the correlation ratios for each axis. The test statistics F_0 is calculated as :

$$F_o = \frac{R^2/p}{(1 - R^2)/(n - p - 1)} \tag{2}$$

where R^2 is correlation ratio, n is sample size, and p is parameter size. All of the ratios are significant at the 1% significance level. Furthermore, the discriminant function correctly classified the five remaining systems.

The validity of our hypothesis is thus demonstrated. Namely, there is a close relation between parameters that reflect user satisfaction and system utilization. It is therefore concluded that the questionnaires shown in Table 1 are applicable for selecting the parameters for service level management.

Table 2. Analysis by Quantification theory type II on availability

item	category	Axis 1			axis 2		
		category score	range	partial correlation coefficient	category score	range	partial correlation coefficient
1	1	-0.97	1.84	0.54	-0.62	0.90	0.31
	2	0.09			0.28		
	3	0.87			0.00		
2	1	1.02	1.21	0.40	1.21	1.69	0.44
	2	0.53			0.83		
	3	-0.17			0.03		
	4	-0.19			-0.48		
3	1	-0.93	1.32	0.61	0.01	0.01	0.01
	2	0.39			0.00		
4	1	-0.37	0.65	0.34	0.13	1.73	0.49
	2	0.28			0.56		
	3	-0.01			-1.17		
5	1	-0.08	0.89	0.33	0.33	1.37	0.47
	2	-0.27			-1.04		
	3	0.62			-0.09		
6	1	-0.03	0.30	0.10	0.06	1.75	0.36
	2	0.14			0.18		
	3	-0.16			-1.57		
7	1	1.25	2.11	0.74	0.57	0.98	0.33
	2	0.44			0.26		
	3	-0.86			-0.41		
8	1	-2.16	2.90	0.76	-0.21	0.78	0.23
	2	0.74			-0.10		
	3	0.14			0.57

Figure 3. Scatter plots for discrimination on availability

Table 3. Analysis by Quantification theory type II on responsiveness

item	category	Axis 1			axis 2		
		category score	range	partial correlation coefficient	category score	range	Partial correlation coefficient
1	1	-0.73	0.99	0.46	0.26	1.50	0.56
	2	0.26			-0.51		
	3	0.16			0.99		
2	1	1.45	2.21	0.70	-0.45	1.75	0.57
	2	1.08			1.30		
	3	0.22			-0.09		
	4	-0.76			-0.36		
3	1	0.60	0.85	0.50	0.78	1.11	0.47
	2	-0.25			-0.33		
4	1	-0.26	0.45	0.28	0.25	1.07	0.40
	2	0.19			0.27		
	3	0.02			-0.80		
5	1	0.08	0.28	0.16	-0.02	0.33	0.12
	2	-0.11			-0.12		
	3	-0.20			0.21		
6	1	-0.01	0.69	0.22	-0.21	3.25	0.60
	2	-0.10			1.08		
	3	0.59			-2.17		
7	1	0.30	0.37	0.15	-0.11	1.23	0.56
	2	0.05			-0.67		
	3	-0.07			0.56		
8	1	-0.23	1.47	0.61	-1.02	2.05	0.51
	2	-0.29			0.04		
	3	1.18			1.03

Figure 4. Scatter plots for discrimination on responsiveness

Table 4. Correlation ratio

	first axis	second axis
Availability	0.6888*	0.596*
Responsiveness	0.734*	0.625*

* indicates significance at $p < 0.01$ level

5. SYSTEM CLASSIFICATION

In Section 4, we divided systems into three groups to determine the service parameters in terms of availability and responsiveness. In this section, we clarify characteristics of each system group by using the analysis results listed in Table 2 and Table 3. The item (i.e., predictor variable) that has the larger value of a range and a partial correlation coefficient highly contributes to classify systems. After paying attention to such items, we sort out categories that have a larger absolute score. The rest of this section explains the characteristics on each system group.

5.1 System Classification in terms of Availability

(1) Group A1: mean time between halt as key parameter

A system in this group has a small absolute score for the first axis and a negative score for the second axis. Thus, a system that belongs to Category 3 of Item 6 (utilization time) tends to be classified to this group. In other words, a system used for a relatively long time without a break fits this

group. Longer utilization leads to a higher risk of a system down when users are connected to the system. Therefore, mean time between halt is more important than mean time to restore.

Furthermore, a system belonging to Category 3 of Item 4 (peak load type) and Category 2 of Item 5 (frequency of use) also tends to go into group A1. Such systems have a characteristic that users can use them at any time and that the date and hour of use are not fixed. This means that users are likely to get the impression that a system whose halt interval is long is more user-friendly than one whose stop time is short when there are two systems of this group with the same availability.

(2) Group A2: mean time to restore as key parameter

A system in this group has positive values for both the first and the second axis. Hence, a system falling into Category 1 of Item 2 (number of users) and Category 1 or 2 of Item 7 (impact on business) is identified as a member of this group. Such a system has a simple function such as information access for relatively fewer users. A typical case of system utilization is that users find out the time when the system is available on any day. It does not bother users very much even if they cannot use the system at a particular time. In this situation, if the service-halt time is not so long, effects on users are small. The mean time to restore is thus the critical parameter.

(3) Group A3: both parameters are important

The gravity of this group is located in the first quadrant in Figure 3. Category 1 of Item 8 (business impact) is the most influential to on this location. This means that the systems in this group have the characteristic that they incur large financial losses when they are down. Since it is necessary to avoid stopping the service as much as possible, the halt interval must be long. In addition, quick restoration is also required even if the service stops. Both parameters should thus be considered.

5.2 System Classification from a Viewpoint of Responsiveness

(1) Group R1: mean response time as key parameter

A system in this group has negative values for both the first and the second axis. Category 4 of Item 2 (number of users) and Category 1 of Item 8 (business impact) indicate this characteristic. The former category shows that the system has many users. This means that mean response time is a common parameter for evaluating responsiveness and is universally accepted. The latter category shows that the system supports mission-critical business operations. Slow response is fatal in the case of a

mission-critical system; namely, the quicker is better. Mean response time is thus an indispensable index because overall speedup is required.

(2) Group R2: maximum response time as key parameter

A system in this group has a large positive value for the first axis and a negative value for the second axis. Hence, a system belonging to Category 1 of Item 2 (number of users), that is, a system for a small user group, is classified into this group. This can be interpreted as follow: If the response happens to be extremely slow, such a bad experience has a strongly negative effect on satisfaction. The ratio of users who had such a bad experience is high in the case of this type of system. In view of this fact, shortening the guaranteed maximum response time is more important than improving mean time for this group.

(3) Group R3: variance as key parameter

This group has a small absolute value for the first axis and a positive value for the second axis. Category 2 of Item 6 (utilization time) represents this group most strongly. A system in this category is used for day-to-day operations. So stable operation is critical in the case that designated people perform the same type of transaction repeatedly. Namely, the lower the variance of response time is, the more users are satisfied. This is because they can predict the response time and establish their pace of work easily. In this case variance is therefore the key parameter.

6. APPLICATION TO SLA/SLM METHODOLOGY

Our SLA/SLM methodology is enhanced by our analysis on the relation between service parameters and system utilization. Figure 5 shows our SLA/SLM methodology. This is the flow until service level is stabilized for the first time. It goes without saying that SLM process does not be terminated at step 6 and that constant improvement activity goes on.

Our analysis improves the second step: mutual understanding between service providers and users. This step is broken down into three works. First, based on the analysis of step 1, important factors for service users are clarified. Second, in terms of the each factor, the key parameters are discussed. Third, both parties recognize their responsibility each other. The mutual understandings of step 2 are followed by negotiation over the service parameters and their targeted values in step 3.

In general, the above second work is executed through interview with service users. However, users tend to insist that everything is important when they are asked a vague question on what is important to their business.

Thus, it is difficult to find out valuable information. We improved this situation by using our analysis. First, we analyze the system utilization from the data gathered in step 1. Second, we estimate the critical service parameters. Then, we confirm with users on our estimation. This approach leads us to effective consensus-building on the important parameters.

Figure 5. SLA/SLM methodology

7. CONCLUSION

The main purpose of the service level management is to increase the user satisfaction. The study outlined in this paper focused attention on system utilization to select the service parameters that reflect user satisfaction. To

investigate parameters for availability and responsiveness, which are known as two major factors in SLA, the relation between service parameters and system utilization was analyzed by Quantification theory type II. The results demonstrated that this relation is strong and enabled us to classify information systems. Therefore, investigation on system utilization can identify the appropriate service parameters for service level management.

A similar approach to that described here will be applied to factors other than availability and responsiveness, such as workload and accuracy. In addition, future works will aim to construct a guideline for determining which factors should be considered for defining SLA/SLM.

REFERENCES

- Adelman, L. et al., 1985. User and R&D Specialist Evaluation of Decision-Support Systems. In *IEEE Trans. on Syst., Man, Cybern.*, vol. SMC-15, No.3, pp.334-342.
- Butler, T.W., 1983. Computer Response Time and User Performance. In *Proc. of ACM CHI'83 Conf. on Human Factors and Computing Systems*, pp.58-62.
- Culwin, F. and Faulkner X., 2001. Browsing the Web: Delay, Determination and Satisfaction. In *Proc. of 34th Hawaii Int'l Conf. System Science*, pp.1802-1806□
- Doll, W. J. and Torkzadeh, G., 1988. The Measurement of End-User Computing Satisfaction. In *MIS Quarterly*, 12, pp. 259-274.
- Endo, Y. and Seltzer, M., 2000. Improving Interactive Performance Using TIPME. In *ACM SIGMETRICS Performance Evaluation Review, Proceedings of the Int'l Conf. on Measurements and modeling of computer systems*, vol.28, Issue 1, pp. 240-251, June.
- Glass, R. L. and Vessey I., 1995. Contemporary Application-Domain Taxonomies. In *IEEE Softw.*, vol.12, pp. 63-76, July.
- Guynes, J. L., 1988. Impact of System Response Time on State Anxiety In *Communications of ACM*, Vol. 31, No. 3, pp.342-347, March.
- Hiles, A., 1999. *Complete Guide to IT Service Level Agreements: Matching Service Quality to Business Needs*, Rothstein Associates Inc.
- Jiang, J.J. and Klein G., 1999. User Evaluation of Information Systems: By System Typology. In *IEEE Trans. on Syst., Man, Cybern. Part A*, vol. 29, No. 1, pp.111-116, Jan.
- Komazawa, T., 1982. *Quantification theory and data processing*, Asakura Books (in Japanese).
- Lewis, L., 1999. *Service Level Management for Enterprise Networks*, Artech House Inc.
- Muller, N. J., 1999. Managing Service Level Agreements. In *Intl. J. Network Management*, 9, pp.155-166.
- Nextslm.org, 2000. Service Level Management – North America Survey 2000. On <http://www.nextslm.org/SLMSurvey2000.pdf>
- Park, J. et al., 2001. Management of Service Level Agreements for Multimedia Internet Service Using Utility Model. In *IEEE Communications Magazine*, pp. 100-106, May.
- Ritchie, I. and Roast, C., 2001. Performance, Usability and the Web. In *Proc. of 34th Hawaii Int'l Conf. on System Science*, pp.1797-1801.