

QoS Assessment of Providers with Complex Behaviours: An Expectation-Based Approach with Confidence

Gareth Shercliff, Jianhua Shao, W. Alex Gray, and Nick J. Fiddian

School of Computer Science, Cardiff University, UK

Abstract. Service Level Agreements (SLAs) define a set of consumer expectations which must be met by a provider if a contract is not to be broken. Since providers will potentially be providing many different services to thousands of different consumers, they must adopt an efficient policy for resource management which differentiates consumers into service ranges. Existing approaches to QoS assessment of providers assume that the policy of a provider with respect to consumers is handled on an individual basis. We maintain that such approaches are ineffective when providers adopt a policy based on service differentiation and in response introduce and evaluate an expectation-based approach to QoS assessment which presupposes the classification of consumers into ranges defined by their expectation. As well as carrying out assessment to determine the likely future behaviour of a provider for a given consumer expectation, we attach a confidence value to our assessment to indicate the level of certainty that the result is accurate. Our results suggest that our confidence-based approach can help consumers make better informed decisions in order to find the providers that best meet their needs.

1 Introduction

Before a consumer and provider enter into an instance of service provision, a set of mutually agreeable criteria must be defined in the form of an SLA[2] in order that both sides are aware of their commitments and of what they should expect from the other party. In such an agreement, the consumer's commitments are usually limited in number and trivial in enactment. Conversely, the commitments of a provider to an individual consumer may be numerous and complex. Consumers may specify QoS parameters such as availability, throughput and response time[9] and there may be interrelationships between particular clauses of the contract. A provider may concurrently be providing several service types, interacting with thousands of consumers and having to efficiently manage resources[11][7] in order that their commitments to each individual SLA are met.

Service differentiation[7] is a generally accepted solution to dealing with the complexity of resource management in flexible domains such as service oriented environments[6]. Taking this approach, consumers are classified into performance classes, where each class represents a set of consumers with particular SLA-defined expectations and commitments. Consumers within the same performance

class can be expected to be treated equally - that is, consumers with similar expectations and commitments will be treated in a similar way by the provider. The policy adopted in classifying and managing consumers in performance classes will differ from provider-to-provider and it can be assumed that providers will not publish details of their policy. Hence, providers are unlikely to advertise each service level as a separate service - to consumers and third-parties it will appear as if all service levels are actually one offering. This presents a challenge to third-party services such as QoS Assessment tools and reputation brokers that attempt to evaluate the likely future performance of a provider, and exposes inadequacies in existing approaches.

Our contribution, which we present in this paper, is an approach to QoS assessment which presupposes that an individual provider may be offering a set of functionally identical services differentiated only by QoS but which are advertised as a single service. We previously introduced an approach to assessment which recognised this differentiation of services[13]. Here we build on our approach by taking into account confidence in the assessment process as an indicator of how likely it is that our assessment is accurate. In this paper we outline our improved approach and demonstrate cases where the confidence in an assessment result may be a discriminating factor in evaluating and choosing between providers.

The remainder of this paper is structured as follows. In Section 2 we briefly discuss existing work similar to our own, identifying points of difference and similarity between the approaches. This is followed in Section 3 by a brief formalisation of the concepts and processes of a model SOC environment that we refer to in our work. In Section 4, we provide a description of our expectation-based approach to assessment, focussing on the incorporation of confidence into our model. In Section 5, we provide evidence of and discuss a set of experiments that we have carried out in order to verify the effectiveness of our approach compared to a non-confidence based equivalent. Finally, in Section 6, we present our conclusions.

2 Related Work

Our work falls into the area of QoS Aware Service Discovery and Selection (QoSADS) the goal of which is to develop techniques to determine the likely future behaviour of a provider based on information that may be provided as ratings[5] or by directly monitoring services[1]. Whilst having the same overall goal, our approach can be distinguished from the majority in taking an expectation-based approach by recognising that the level of service delivered by a provider is partially dependent upon the expectation of the consumer. We previously developed such an approach based upon consumer ratings[4]. An expectation-based approach is also adopted by Scherchan et al.[12] who concur with our assertion in using similar expectation as a discriminating factor in selecting relevant past provision instances for use in assessment. However, we maintain that when delivered values are used instead of ratings, simply choosing past instances of provision where expectation is similar is ineffective when providers adopt a differentiated services approach to service provision[13].

The research into trust and reputation has developed techniques to choose the best provider to meet a consumer's needs, though the emphasis in this case is on *how likely* the provider will be to keep to contractually agreed levels of service[10], rather than determining the actual level of service that can be expected. In [8] Maximilien and Singh provide a comprehensive approach to establishing reputation of providers taking into account both rated and recorded levels of service.

3 Problem Statement

We here provide a brief formalisation of the concepts and processes of a service-oriented environment relevant to our work, concluding the section with a statement of the problem addressed by our approach.

3.1 Quality

We adhere to the conformance view of quality, used commonly in the literature[5][3] - specifically, we take the approach that conformance is the act of meeting consumer expectation.

Definition 1 (Quality). *Let Q be the degree of conformance of the service delivered (D) by a service provider to a consumer's expectation (E).*

We assume that the definition of quality above is used rationally by both consumers and providers in their environment. That is, consumers are happier when their expectations are being met and providers believe that they are able to keep consumers happy by minimising the distance between the level of service they are providing and the consumer's expectation.

3.2 Service Provider Behaviour

We assume a problem space within an SOC environment, in which a set of service providers $PSet = \{P_1, \dots, P_n\}$ offer functionally identical services. Each individual service provider will adopt a policy to resource management by classifying consumers based on their expectation - effectively dividing the range of consumer expectation into partitions such that the whole range of expectation is covered. We refer to each partition as a *service range* (SR). The ranges are defined by an upper and lower expectation - all consumers with expectation falling into a particular range will be treated similarly by the provider.

Within each service range, at a given point in time, we assume that a provider has a target level of service that they attempt to provide (denoted D_t) though generally they will not be able to maintain performance at exactly this level. We represent the actual level of service delivered at any point as consisting of D_t and an error, ϵ . D_t may change over time dependent on other factors such as a change in the amount of resources allocated by a provider to a particular service range.

Definition 2 (Target Service Level). Let D_t be the level of service which a provider attempts to meet for all consumers classified within a single expectation-delimited service range and ε be the difference between the level of service received by the consumer at any point and D_t .

It is important to reiterate at this point that explicit knowledge about a provider's target service levels or the extent of the service ranges remains unavailable to anyone except the provider.

3.3 Collection and Utilisation of QoS Information

When a service is provided to a consumer, we assume that data about the instance is collected. We refer to all QoS information about a single instance as a quality datum QD, and the set of all quality datum as the quality database (QDB):

$$QDB = \{QD_0, \dots, QD_p\}$$

$$QD_i = (time_i, P_i, exp_i, del_i)$$

Where $time_i$ was the time that the service instance terminated, P_i was the provider that provided the service (from PSet), exp_i was the consumer's expectation and del_i was the average level of service provided (measured on the same scale as exp_i). For simplicity and ease of presentation, we assume that exp and del are normalised onto a scale of (0..1).

Definition 3 (Quality Assessment). Given a consumer expectation E_c , a provider $P \in PSet$ and the contents of QDB, the goal of QoS assessment is to determine the likely behaviour of P if he were to agree to provide E_c .

In carrying out such an assessment, it is necessary to consider both the effect of differing expectation of consumers - two individuals with different expectations may gain a different utility from the same delivered level of service (from Definition 1); and the behaviour of individual providers in relation to a consumer's expectation - providers will adopt different policies with respect to defining and adhering to service levels and these are not explicitly available to the assessment service (from Definition 2).

3.4 Notation

We can consider the data from the QDB regarding a single provider as being a set of tuples (e,d). In order to illustrate our approach, we utilise a graphical notation (Figure 1) in which we define two interconnected spaces: e-space and d-space; containing tuple data with a one-to-one mapping between spaces. That is, each tuple corresponds to exactly one point in each space.

4 An Expectation-Based Approach with Confidence

In this section we describe our approach to QoS assessment. A more comprehensive overview of our initial approach is provided in [13]. Here, we concentrate on identifying the factors which will affect the confidence of our assessment,

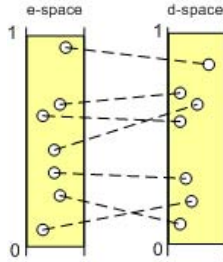


Fig. 1. Quality Spaces: e-space and d-space

formalising their calculation and describing how they are used to improve the effectiveness of our approach.

We consider our approach as consisting of four main phases (Figure 2) which we describe below.

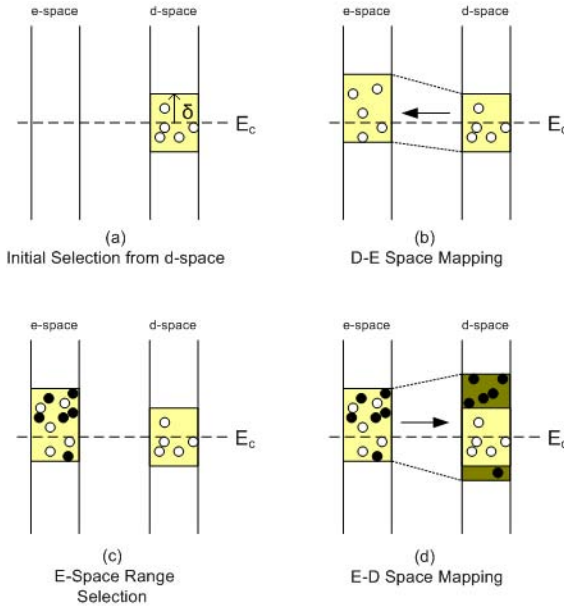


Fig. 2. Stages of the Approach - Mapping, Selection and Aggregation

Initial Selection from D-Space (Figure 2(a)). In order to ascertain whether it is possible for a provider (P_i) to meet the consumer’s expectation (without constraining the circumstances under which this might happen), we attempt to find past instances in which the delivered level of service is similar to that of the consumer’s expectation. We define similarity through the use of a threshold value δ - the range delimited by E_c and δ specifying a range of provided level

of service for which we assume the consumer will be satisfied (with E_c at the centre of the range).

$$ProviderData = \{pd \mid pd \in QDB, pd.prov = P_i\}$$

$$SimilarInstances = \{sim \mid sim \in ProviderData, E_c - \delta \leq sim.del < E_c + \delta\}$$

If any similar instances are found this is an indication that the provider is capable (or has been capable) of providing service at the level desired by the consumer. However, the instances of provision which have been selected may have been provided as a result of a consumer having a different expectation to E_c - for instance, if a provider consistently over- or under-provides. In the next step, we identify whether this is the case.

D-Space to E-Space Mapping (Figure 2(b)). In order to determine the level of consumer expectation which results in E_c being provided, we identify the corresponding e-space points for each element of *SimilarInstances*, determining the lower and upper range of expectation covered by the points - denoted as *lowerExp* and *upperExp* respectively.

If E_c is in the range delimited by *lowerExp* and *upperExp* this should increase our confidence that the provider is capable of delivering a level of service similar to E_c when asked to do so. i.e. the provider has a high degree of conformance when asked for E_c . At this stage, however, we must still consider the fact that the provider's level of service is not constant over time. That is, for other instances with consumer expectation in the range identified, the provider may have delivered different values in the past - if this were the case, it would lower our confidence in our provider's ability to maintain a consistent behaviour at the level E_c .

In general, the confidence that the points identified in e-space are relevant to our assessment is inversely proportional to the distance of E_c from the mean of the expectation identified.

$$conf_{exp} = 1 - |E_c - \mu_{exp}|$$

E-Space Selection (Figure 2(c)). We now select *all* points in e-space where the consumer's expectation in each instance falls into the range delimited by *lowerExp* and *upperExp*. The points identified now cover all data that is potentially relevant to the result of our assessment.

$$PotentiallyRelevant = \{ran \mid ran \in ProviderData, lowerExp \leq ran.exp < upperExp\}$$

E-Space to D-Space Mapping (Figure 2(d)). Finally, we aggregate the values from the corresponding points from d-space in order to determine the predicted behaviour of the provider. As indicated above, the set of data used to carry out the final aggregation and thus predict the performance of the provider contains all past instances which are *potentially* relevant. In an ideal situation,

the points used in the final aggregation will fall into a single service range of the provider (Figure 3 (a)). If this is the case, by aggregating the delivered values we should obtain a prediction which falls into the same range and is thus a good indication of performance. However, in other situations the instances which we identify may fall across multiple service ranges (Figure 3 (b)). In this case, aggregating the delivered values will produce a prediction which falls between the ranges identified and is thus unlikely to be accurate.

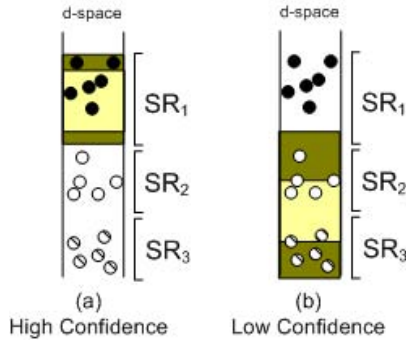


Fig. 3. Confidence in data used for final aggregation

To quantify the level of confidence we have in the data we consider the likely distribution of data within each of the service ranges. As discussed in Section 3.2, a provider will attempt to meet a target level D_t but in each instance will generally miss this objective by an error defined as ϵ . It is reasonable to assume that in each case, the probability of ϵ being small is greater than the probability of ϵ being large - that is, in general the provider is likely to miss D_t by a small amount more often than a large amount. We can therefore consider the distribution of recorded delivered values in each service range as being normal, with D_t as the mean.

The confidence which we can have in the data is proportional to the probability that the data we have found falls into a single service range. We use likelihood estimation to determine the probability of the observed data being generated as the result of a normal distribution (whose mean and variance are derived from the data). The higher the likelihood, the more likely that our data falls into a single range.

$$\mu = \frac{\sum_{i=1}^n del_i}{n} \tag{1}$$

$$\sigma^2 = \frac{\sum_{i=1}^n del_i^2 - (\sum_{i=1}^n del_i)^2/n}{n - 1} \tag{2}$$

$$conf_{del} = P(RelevantInstances|M_{\mu,\sigma}) = \prod_{i=1}^n P(relevantinstances_i.del|M_{\mu,\sigma}) \tag{3}$$

Where M is a model of a normal distribution defined by μ and σ .

Overall Confidence and Result Calculation. The overall confidence in assessment is defined as the product of the two separate confidences from d-space and e-space. The predicted value of performance for the provider is taken as the mean of the final set of points identified in d-space.

$$conf = conf_{exp} * conf_{del} \quad (4)$$

$$prediction = \frac{\sum_{i=1}^n del_i}{n} \quad (5)$$

Once overall confidence and the predicted performance for the provider have been determined, both values are passed back to the consumer.

5 Evaluation

We have created a discrete-time event simulator which allows fine-grained control over the parameters of a model SOC environment described in Section 3. The simulator allows the modelling of provider behaviour in terms of service levels and QoS delivered over time; service discovery and provision in which expectation and delivered values are recorded in a QDB; and provides a scripting language for the description and simulation of specific scenarios. We used this simulator in order to carry out a number of experiments in order to validate our assertions and to verify the effectiveness of our approach.

5.1 Empirical Results

In the first experiment, we observed the behaviour of our approach in terms of how well the performance of a single provider could be predicted. Figure 4 illustrates the behaviour of a single provider over time. Here, the provider's policy divides the expectation space into two ranges - the provider giving a level of service of around 0.55 for $exp > 0.35$ and a level of service of around 0.4 for $exp < 0.35$. The performance of the approach can be evaluated by observing how far the predicted value falls from the actual delivered level of service. The bars for each point on the graph indicate the confidence in the result - the larger the bar, the higher the confidence.

In our second experiment, we compared the performance of non-confidence and confidence based approaches in selecting between two providers. In this case, the graphs (Figure 5) illustrate the behaviour of each provider for the consumer's expectation. For each approach, we plot a point at each assessment during the simulation to indicate which provider was chosen by each algorithm. At any point in the simulation, the best provider (and hence the one that should be selected by the algorithm) is the one whose delivered level of service is closest to E_c .

Experiment 1 - Confidence as an Indicator of Uncertainty. We defined a provider behaviour in which the delivered values for each service range both intersect the range defined by $E_c \pm \delta$ i.e. from 0.4 - 0.6. In this case, the actual delivered level of service which would be received for E_c is defined by the upper

service range (since $E_c > 0.35$). However, in this instance we are providing the assessment algorithm with a behaviour in which the delivered values for both service ranges overlap the range defined by $E_c \pm \delta$. We are therefore concerned with both how the effectiveness of prediction is affected and the level of confidence that the algorithm has in its assessment. The results of this experiment are illustrated in Figure 4.

Figure 4 shows that initially the confidence-based algorithm predicts a default value for the provider (this is to be expected as the algorithm has not yet

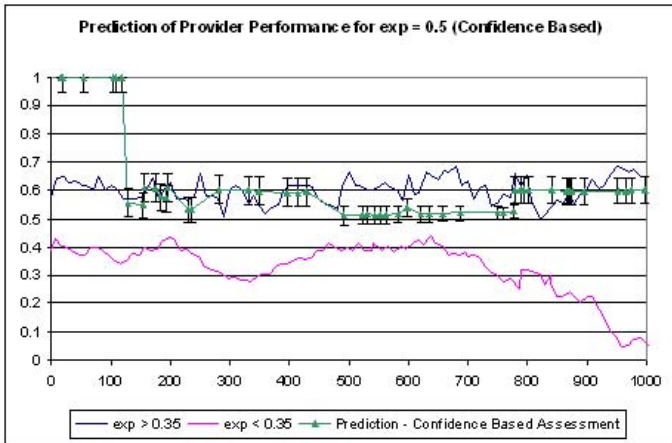


Fig. 4. Experiment 1 - Confidence as an Indicator of Uncertainty

observed any provision in the range defined by $E_c \pm \delta$). From $t=100$ to $t=400$, the algorithm correctly predicts performance in accordance with the upper service range. However, between $t=450$ and $t=750$ the level of service delivered by the lower service range falls into the range $E_c \pm \delta$. In this case, the assessment algorithm predicts a level of service between the two service ranges, though the confidence in the assessment is now substantially reduced. As the delivered level of service for the lower service range drops towards $t=1000$, both the predictive accuracy and confidence of the assessment return to normal. This experiment illustrates that the inclusion of confidence within the assessment has the desired effect. When the predictive performance of the approach is affected by an overlap in the performance of the provider’s service levels, confidence in the assessment falls appropriately.

Experiment 2 - Choosing between multiple providers. In our second experiment, we defined behaviour for two providers. Provider 1 has a single behaviour which begins providing service at 0.6, but drops to about 0.4 (E_c in this case) later in the simulation. Provider 2 has multiple service ranges - in this simulation, the level of service offered to consumers requesting $E_c = 0.4$

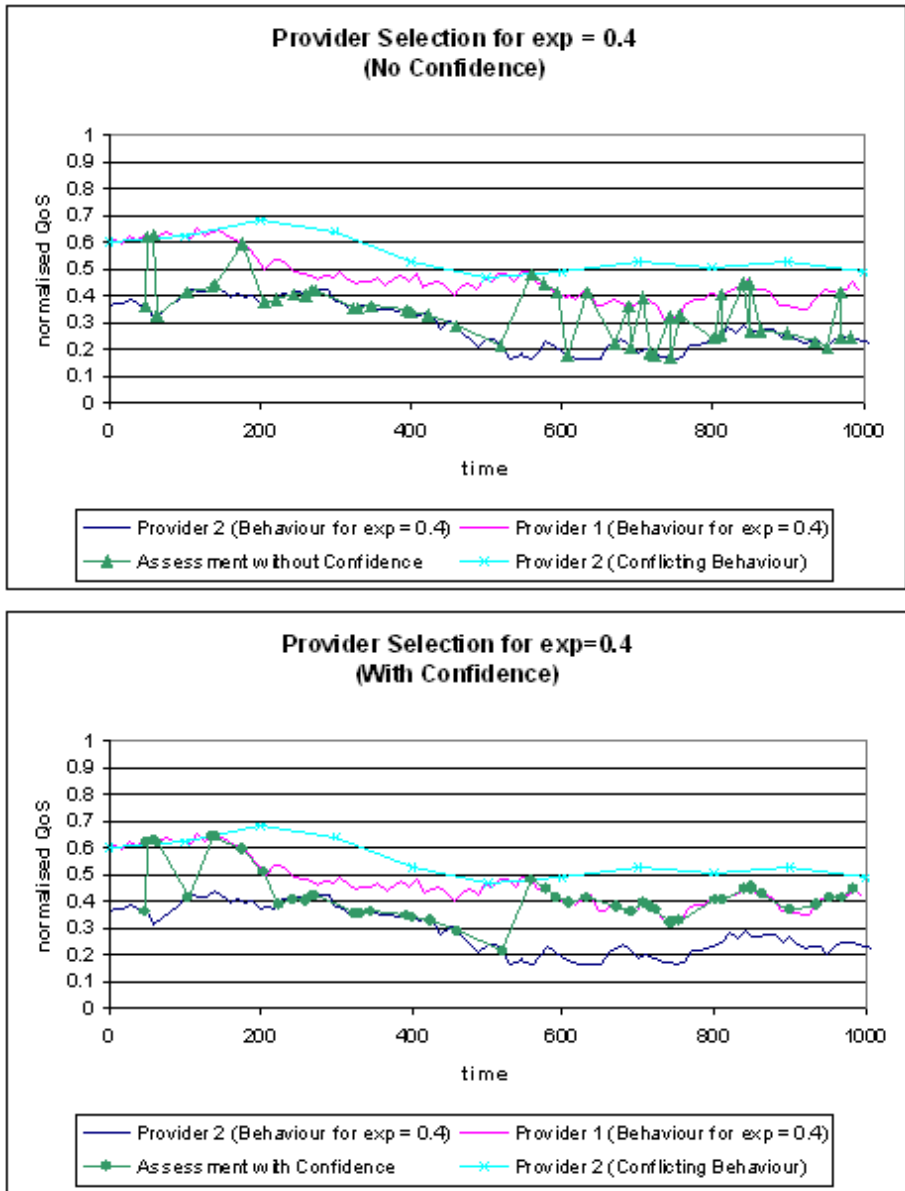


Fig. 5. Experiment 2 - Choosing between Multiple Providers

begins at 0.4, but falls to 0.2 later in the simulation. The second behaviour of Provider 2 would be received from the provider for $E_c \neq 0.4$, and as such should be irrelevant to our assessment.

The behaviour observed in Figure 5 can be described in terms of three phases. In the first phase ($t=0$ to $t=200$) both algorithms alternate between Provider 1 and Provider 2. This is due to the fact that the information available to the algorithms in making their assessments is not sufficient in order to adequately evaluate each provider so each is given a default assessment and a random provider is selected. From $t=200$ to $t=450$, both algorithms assess Provider 2 as being the provider that will closest meet E_c . At the beginning of this phase, this is certainly the case. However, by the end of the phase the performance of Provider 2 has decreased such that it is now significantly poorer than Provider 1 at meeting E_c . During the final phase ($t=500$ to $t=1000$) the non-confidence based approach begins to alternate between assessing Provider 1 and Provider 2 as the best provider. This is due to the phenomenon observed in Experiment 2 - the range defined by $E_c \pm \delta$ overlaps with both service ranges of Provider 2 and thus provides an inaccurate assessment, which in this case predicts a level of service very close to that of Provider 1. Conversely, although the confidence-based approach will predict the same level of service as the non-confidence based approach, it will be less confident in Provider 2's capability than Provider 1's. This can be observed in the final phase as the confidence-based approach consistently recommends Provider 1 as the provider who is closest to E_c . This reaffirms the effects of the results already observed in Experiment 1 and illustrates the significance of including confidence as part of an expectation-based approach to assessing and choosing between multiple providers in an SOC environment.

6 Conclusion

In this paper we have presented an improved approach to expectation-based QoS assessment which attaches a confidence indicator to each assessment result indicating a degree of certainty in its accuracy. The derivation of the confidence measure is based on a combination of the relevance of the range of expectation identified to that of the consumer; and the likelihood of the set of data identified as relevant in making the assessment corresponding to a service range defined by the provider's resource management policy. We have demonstrated using experimental evidence that by adding confidence to the assessment process, consumers are able to make a better informed decision of which provider to select, thus increasing their overall utility. In future work, we intend to build upon our approach by using clustering techniques in order to explicitly identify the service ranges offered by a particular provider.

References

1. Rashid J. Al-Ali, Omer F. Rana, and David W. Walker. G-QoSM: Grid Service Discovery Using QoS Properties. *Journal of Computing and Informatics*, 2003.
2. A. Dan, D. Davis, R. Kearney, R. King, A. Keller, D. Kuebler, H. Ludwig, M. Polan, M. Spreitzer, and A. Youssef. Web Services on Demand: WSLA-Driven Automated Management. *IBM Systems Journal, Special Issue on Utility Computing*, 43(1):136–158, March 2004.

3. V. Deora, J. Shao, W.A. Gray, and N.J. Fiddian. A Quality of Service Management Framework Based on User Expectations. In *Lecture Notes in Computer Science*, volume 2910, pages 104–114, January 2003.
4. Vikas Deora, Jianhua Shao, W. Alex Gray, and Nick J. Fiddian. Expectation based quality of service assessment. *to appear in International Journal on Digital Libraries*, 2006.
5. Sravanthi Kalepu, Shonali Krishnaswamy, and Seng Wai Loke. Reputation = f(User Ranking, Compliance, Verity). In *Proceedings of the IEEE International Conference on Web Services (ICWS'04)*. IEEE Computer Society, 2002.
6. Ronald M. Levy, Jay Nagarajarao, Giovanni Pacifici, Mike Spreitzer, Asser N. Tantawi, and Alaa Youssef. Performance Management for Cluster Based Web Services. In Germán S. Goldszmidt and Jürgen Schönwälder, editors, *Integrated Network Management*, volume 246 of *IFIP Conference Proceedings*, pages 247–261. Kluwer, 2003.
7. Heiko Ludwig. Web Services QoS: External SLAs and Internal Policies or: How do we deliver what we promise? In *Fourth International Conference on Web Information Systems Engineering Workshops (WISEW'03)*, pages 115–120, 2003.
8. E. Michael Maximilien and Munindar P. Singh. Toward Autonomic Web Services Trust and Selection. In *ICSOC '04: Proceedings of the 2nd International Conference on Service Oriented Computing*, pages 212–221. ACM Press, 2004.
9. Daniel A. Menasce. QoS Issues in Web Services. *IEEE Internet Computing*, 6(6):72–75, November 2002.
10. Sarvapali D. Ramchurn, Dong Hunyh, and Nicholas R. Jennings. Trust in Multi-Agent Systems. *Knowledge Engineering Review*, 2004.
11. A. Sahai, J. Ouyang, V. Machiraju, and K. Wurster. BizQoS: Specifying and Guaranteeing Quality of Service for Web Services through Real Time Measurement and Adaptive Control. HPL-2001-134. Technical report, HP Labs, 2001.
12. Wanita Sherchan, Shonali Krishnaswamy, and Seng Wai Loke. Relevant Past Performance for Selecting Web Services. In *QSIC '05: Proceedings of the Fifth International Conference on Quality Software*, pages 493–445, Washington, DC, USA, 2005. IEEE Computer Society.
13. Gareth Shercliff, Jianhua Shao, W. Alex Gray, and Nick J. Fiddian. A Multiple Quality-Space Mapping Approach to Qos. In *CIT 2006: Proceedings of the Sixth IEEE International Conference on Computer and Information Technology [to appear]*. IEEE Computer Society, September 2006.