5-2013

Modelling Confidence for Quality of Service Assessment in Cloud Computing

Hmood Al-Dossari
Department of Computer Science King Saud University Riyadh, Saudi Arabia, hzaldossari@ksu.edu.sa

Jianhua Shao
School of Computer Science and Informatics Cardiff University, j.shao@cs.cardiff.ac.uk

Modelling Confidence for Quality of Service Assessment in Cloud Computing

Hmood Al-Dossari
Department of Computer Science
King Saud University
Riyadh, Saudi Arabia
hzaldossari@ksu.edu.sa

Jianhua Shao
School of Computer Science and Informatics
Cardiff University
Cardiff CF24 3AA, UK
j.shao@cs.cardiff.ac.uk

Abstract
The ability to assess the quality of a service (QoS) is important to the emerging cloud computing paradigm. When many cloud service providers exist offering many functionally identical services, the prospective users of these services will wish to use one that offers the best quality. Many techniques and tools have been proposed to assess QoS, and the ability to deal with uncertainty surrounding the QoS verdicts given by any such techniques or tools is essential. In this paper, we present a probabilistic model to quantify confidence in QoS assessment. More specifically, we take the number of QoS data items used in assessment and the variation of data in the dataset into account in our measure of assessment reliability. Our experiments show that our confidence model can help consumers to select services based on their requirements effectively.

Keywords
QoS Assessment, Confidence, Data Size, Data Deviation, Service Selection.

1. Introduction
Quality of service (QoS) assessment may be broadly described as a function to determine, using historical service provision data, the likely quality of a service that a consumer may get from a provider. The ability to assess the quality of a service is important to the emerging cloud computing paradigm. When many cloud service providers exist offering many functionally identical services, the prospective users of these services will wish to use one that offers the best quality. So techniques and tools that can help assess QoS are desirable for the consumers.

In a cloud computing environment, any techniques and tools that attempt to help assess QoS will have to deal with some uncertainty surrounding the QoS verdicts they give. This is because in such an environment, service providers' behaviours cannot always be expected to be stable, and their performance may fluctuate due to a range of factors, e.g. network congestion, resource constraints or simply lack of good quality management (Sahi. 2001, Kim. 2005, Jureta. 2009, Wohlstadter. 2004). It is easy to see that such variation in service delivery performance will be reflected in the collected QoS data, and a QoS verdict based on a set of data that has a large variation is likely to be unreliable. Moreover, when QoS assessment is conducted on a small set of data (e.g. when assessing a new service), then intuitively the verdict may not be reliable either. Thus, it is essential that we attach a confidence value to a QoS assessment, so that the level of certainty surrounding the verdict can be indicated to the assessment requester, and can be taken into account when he or she selects the preferred service to use.

In this paper, we present a probabilistic model to quantify confidence in QoS assessment. We do so by integrating two reliability measures: the number of QoS data items used in assessment and the variation of data in the dataset. We show that our confidence model can help consumers to select services based on their requirements effectively.

The remainder of the paper is organised as follows. Section 2 describes how decision theory may help address uncertainty that consumers may face in making a decision about which service to use. Section 3 presents related work. Section 4 provides a description of our proposed model to deal with uncertainty surrounding QoS assessment. We report experimental results in Section 5 and conclude the paper in Section 6.

2. Dealing with uncertainty

To deal with uncertainty in QoS assessment, we use Decision theory (Svenson, 1979). The key idea underlying decision theory is quite simple: people always choose actions that move them towards situations, or states, that they prefer. Given a set of alternatives and a set of consequences following each alternative, decision theory models the relationship between the two sets and offers a conceptually simple procedure for choosing among alternatives. For example, if Alice wants a web hosting service that can serve 800 requests per second, and she needs to choose between two web hosting services S1 and S2. Suppose that S1 is able to deliver 200 requests per second, whereas S2 can deliver 800 requests per second. Then, Alice would prefer S2 to S1, as taking this action will leave her in a better 'state', i.e. receiving a throughput of 800 requests per second.

To capture what is a preferred state to a consumer, consumer expectations (i.e. requirements) on quality of services are encoded in a utility function that maps the set of situations that the consumer may find himself into a set of real numbers representing the value of each situation to that consumer. To move to the state that they most prefer, a consumer always chooses actions that maximise his utility. This is done such that, if one state is preferred to another, then the utility value returned for the preferred state will be higher than that for the other. Similarly, if two states are equally preferred, then they should have equal utility. Formally, if Θ is the set of all states the consumer can reach through his own actions, and the utility of each $\theta \in \Theta$ is given by the utility function $U(\theta)$, then the consumer should choose to act so as to arrive in a state φ, such that:

$$\forall \theta \in \Theta, U(\theta) \leq U(\varphi)$$

Unfortunately, it is not always feasible to know how to act in order to arrive in state φ. Typically, if a consumer has to choose between competing providers (such as S1 and S2 in Alice's example) he will not know for certain which provider will act most in his favour (i.e. be closest to meeting his expectation). To deal with this problem, decision theory draws upon probability theory (Milton, 2002) and provides a reasonable method to address this uncertainty. That is, the consumer should act to maximise expected utility, i.e. the utility they expect to obtain from a service if the provider is true to what he has promised, and the likelihood that the provider is able to do this. More formally, expected utility function for a single attribute can be defined as follows:

Definition 1 (Expected Utility Function for Single Attribute) The consumer C_j's expected utility function for attribute A_k of service S_i is defined as:

$$EU^j_k(\theta) = \int_{\theta \in \Theta} p(\theta|\alpha)U(\theta)d\theta$$

where θ is the state that C_j will move to w.r.t. attribute k and $p(\theta|\alpha)$ is the probability of θ given that the consumer takes action α, i.e. selects to use service S_i.

To derive a representation of consumer's utility $U(\theta)$ in Definition 1, we need to understand consumers' expectations or requirements for QoS. Conformance has been defined in (Deora, 2003) as the means by which consumers judge the quality of a service: the quality is deemed to be maximised when the consumer's expectation is met exactly by the service provider. So, the utility ($U(\theta)$) obtained from a particular service will correspond to the degree of conformance to the consumer's expectation.

The presence of probability in the definition of expected utility means that decision theory in the face of uncertainty enters the realm of statistics. That is, to determine the expected utility of an action, we must determine the probability distribution of the possible states in which an action will result. How we determine this distribution depends on the nature of the problem at hand. The decision of choosing between *n* candidate services based on information from a QoS assessment method can be viewed as choosing a service that exhibits the best expected utility. In such a decision, the factor that potentially affects the utility of each choice is the accuracy of prediction by a QoS assessment method regarding the performance of the candidate service. Calculating the expected utility as in Definition 1 for each candidate service will result in an expected utility value for each choice. Rationally, a consumer will choose the candidate service that corresponds to the action offering the highest expected utility value. More specifically,

$$S_{opt} = \arg\max EU(S_i) \qquad \textbf{Equation (1)}$$

It should be noted that the decision made by the consumer based on Equation 1 considers QoS only. There are many other factors that a consumer may consider when choosing a service, for example, the cost or reputation of a service provider. To accommodate these factors in consumers' decision-making, the modeling of expected utility could be extended. In this paper, however, we do not consider this issue.

3. Related Work

There are many confidence models proposed in the literature, especially in the area of trust and reputation systems. The main aim of these models is to deal with uncertainty surrounding the behaviour of consumers when producing their ratings. Sabater et al. (Sabater et al. 2001) introduced two measures to calculate the reliability of trust value; the number of ratings and their deviation. The authors in (Huynh, 2004) and (Keung, 2008) concur with the assertion of using these two measures to calculate trust value's reliability. The work in (Mui, 2002) uses the Chernoff Bound to determine the minimum sample size required to achieve a certain level of confidence. This approach is also used by the author in (Zhang, 2006) to compute if the experience (measured by sample size) of an agent is sufficient enough (i.e. reliable enough) to reason about the likely behaviour of other agents. Overall, these studies have noted the importance of considering data characteristics (measured by size and deviation) in deriving a confidence value, but these confidence models cannot be directly applied in our work for two reasons. First, they are more for discrete and subjective data (user ratings), while the data we are dealing with is continuous and objective (monitored data). Second, the reliability verdict is derived "statically" and ignores consumer expectations, which is important in calculating our utility.

Adopting the conformance view of quality imposes more complexity in confidence computation. This is because what is assumed to be good enough for one consumer may not be for another. For example, a hosting web service that delivers 400 requests per second to Alice who was promised to receive 800 requests per second will be judged to be unsuccessful. However, this does not mean that the same delivered value (400 requests/second) will be judged by other consumers in a similar way. For example, the same delivered level of service would be enough and satisfactory for someone who expected 390 requests per second. So, in general, in contrast to the existing confidence methods, to determine whether a given data instance is reliable or not we must "dynamically" derive a verdict with respect to the required quality. In the next section, we propose a confidence model that can handle this issue adequately and deal with uncertainty surrounding a QoS prediction.

4. Proposed Confidence Model

When a QoS verdict is derived for a given service using its past performance data, it is essential that we are able to establish its reliability or place a confidence on the verdict. In our context, the confidence is used to deal with uncertainty surrounding the reliability of a QoS prediction. More specifically, it expresses how confident QoS assessment method is in producing the assessment result given the data used in assessment. This is an important measure because a consumer's decision about which service to choose may depend on the service's behaviour over time, e.g. its stability to deliver the required level of service. So a good QoS assessment should give an indication on the reliability of its assessment result, so that consumers will be able to make a better informed decision in selecting their preferred services.

In our confidence model, we use two measures to calculate the confidence of a prediction for a single attribute: data size (Rel_{ω}) and data deviation (Rel_{θ}). The former indicates how strongly the prediction derived by the QoS assessment method is supported by the dataset, and the latter indicates the service's consistency in delivering that prediction. In the following sections, we establish these two measures formally, based on the conformance view of quality as explained in Section 2.

4.1 Data Size Measure (Rel_ω)

This measure is based on the number of data items used in assessment. Each data item used in assessment provides an independent piece of evidence about the quality that a service has offered in the past. So intuitively, the more evidence we have, the higher confidence we should have for the assessment. More formally, this is captured in Equation 2 which states that as the number of data items grows, the degree of reliability increases until it reaches a defined threshold denoted by m:

$$Rel_\omega = \begin{cases} \frac{n}{m} & \text{when } \bar{n} < m \\ 1 & \bar{n} \geq m \end{cases} \qquad \textbf{Equation (2)}$$

where $\bar{n}$ is the number of data items selected and used in assessment. So Rel_ω increases from 0 to 1 as the number of selected data items $\bar{n}$ increases from 0 to m, and stays at 1 when $\bar{n}$ exceeds m.

To determine the minimum number of data items (m_{min}) needed in order to achieve a certain level of confident about an assessment, we use the Chernoff Bound (Mui, 2002) to calculate m_{min}:

$$m_{min} = \frac{1}{2\varepsilon^2} \ln \frac{1-\lambda}{2} \qquad \textbf{Equation (3)}$$

where ε is the maximal level of error that can be accepted by the consumer, and λ is the required confidence level. So the larger the λ and the smaller the ε are, the larger m_{min} is required. For example, if we set $\lambda = 0.99$ and $\varepsilon = 0.1$, then the minimum number of data items needed is m_{min} = 1060. While this suggests that we should use as many recorded QoS data items as possible in order to have confidence in assessment, care must be taken, as argued by (Al-Dossari, 2010) and (Al-Dossari, 2012a), that we do not use data that may give us a misleading verdict.

4.2 Data Deviation Measure (Rel_θ)

This measure is based on the variation within the data used in assessment. When selected QoS data are aggregated into a single verdict, it is important to take data variation into account. This is because different data distributions may average to the same mean, yet they have different variations (i.e. some may have a more fluctuate service delivery than another). This suggests that using the derived mean alone may not be a reliable verdict.

To capture variation within the data used in assessment, we view service delivery as a set of Bernoulli trials: *successful delivery* (delivered the required service level) or *unsuccessful delivery* (did not deliver the required level), and then model it as Beta distributions:

$$Rel_\theta = \frac{\alpha}{\alpha + \beta} \qquad \textbf{Equation (4)}$$

where α = r + 1 and β = s + 1, and r is the observed number of successful deliveries and s the unsuccessful ones, following Laplace's rule of succession (Ristad, 1995). The ratio of α and β will determine where in the interval [0, 1] the distribution peaks, and a high α will cause the distribution mode to occur close to 1.

To determine whether a past delivery on a single QoS attribute was a successful one or not, we use the following:

$$x(A_i) = \begin{cases} 1 & d(A_i) \in [p_i - \vartheta, p_i + \vartheta] \\ 0 & \text{otherwise} \end{cases}$$

where $d(A_i)$ is a delivered quality on A_i in the selected dataset D and p_i is the verdict given by the assessment on attribute A_i. If $d(A_i)$ is within a specified ϑ from p_i, then it is considered to be a successful delivery, denoted by $x(A_i) = 1$. Otherwise, it is unsuccessful, denoted by $x(A_i) = 0$. We call the range, i.e. $[p_i - \vartheta, p_i + \vartheta]$, a confidence range. Accordingly, r and s are:

$$r = \sum (x_i = 1) \text{ and } s = \sum (x_i = 0) \qquad \textbf{Equation (5)}$$

Clearly, choosing different values for ϑ can have a direct impact on Rel_θ. For the same data size, a large ϑ can result in a more confident QoS verdict, because it can result in more deliveries that will be considered to be successful and consequently can increase Rel_θ. Thus, a large ϑ may be chosen when a consumer is willing to accept a more fluctuated delivery of a service from a service provider.

Note that in contrast to other models in the literature (Zhang. 2006, Teacy. 2006, Mui. 2002, Jøsang. 2002), whether a past delivery is a successful one or not is determined dynamically in our model. This is because p_i is computed based on consumer expectation, so whether $d(A_i)$ is a successful delivery or not is affected by consumer expectation or requirement. For example, if *S1* is delivered with a quality of 0.5 on attribute A_i, then it is deemed to be a successful one for a consumer who requested 0.5 on A_i, but not for the one who requested 0.9, if we set $\vartheta = 0.1$.

4.3 Overall Confidence Value

Intuitively, the Rel_ω measure indicates how strongly the mean derived by the QoS assessment method is supported by the dataset and the Rel_θ measure indicates the service's consistency in delivering that mean. Based on these two measures, we calculate the overall confidence as follows:

$$\text{Confidence} = Rel_{\omega} \times Rel_{\theta} \qquad \textbf{Equation (6)}$$

5. Evaluation and Results

We have carried out simulations to study the performance of our proposed confidence model. In our experiments, we simulated four services (S1, S2, S3 and S4), and the test data for the four services were generated according to Table 1.

Table 1: Data Generation for Services

Service	Delivered values	Comment
S1	N (0.5,0.20)	Fluctuating a lot (1000 data tuples)
S2	N(0.5,0.05) then N(0.5,0.20)	Consistent (first 500 data tuples) then fluctuated (second 500 data tuples)
S	N(0.5,0.05)	Consistent with short history (500 data tuples)
S	N(0.5,0.05)	Consistent with long history (1000 data tuples)

The values simulating the delivered quality for each service were normally distributed with the same mean quality (μ = 0.5). However, different sizes and fluctuations, controlled by standard deviation (σ), were used to create the four services to simulate different behaviours. Some example test data are shown in Figure 1.

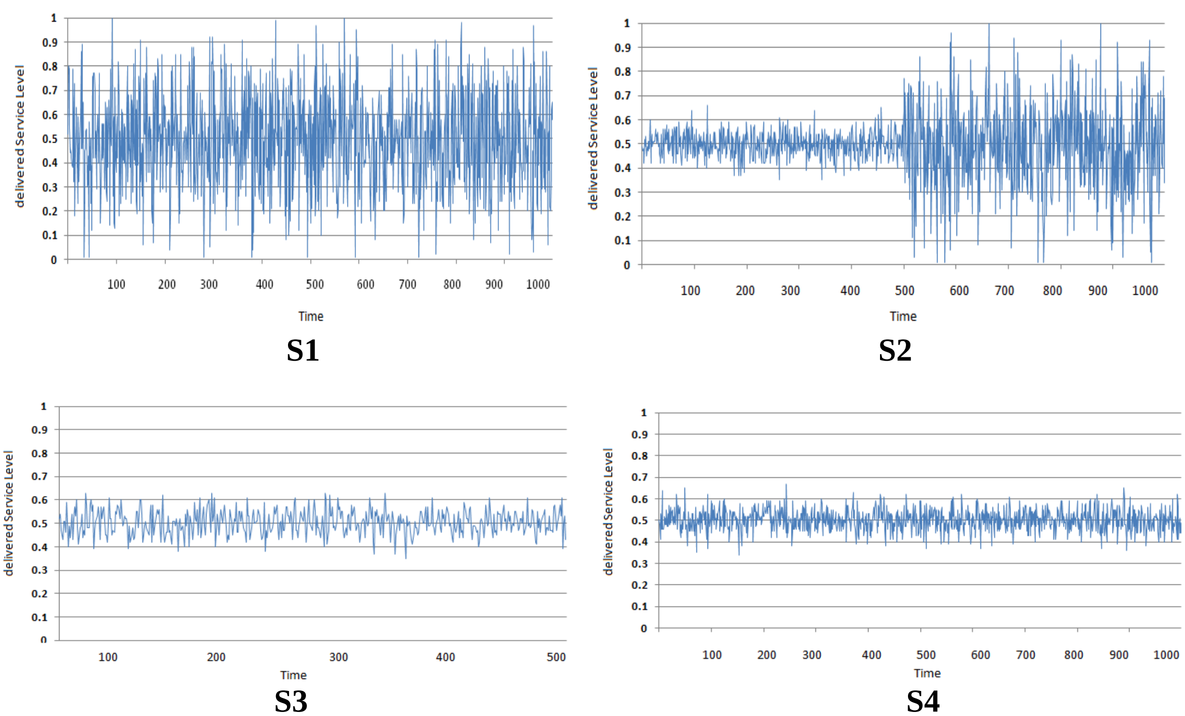

Figure 1: Delivered qualities by the four services

Experiment 1 - Confidence as an Indicator of Uncertainty.

In this experiment, we considered confidence calculation based on the two proposed reliability measures (Rel_ω and Rel_θ). In calculation of Rel_ω, we set the required confidence level to be λ = 0.95 and maximum tolerable error level to be ε = 0.05 (i.e. the minimum number of data items needed to be confident about the assessment result m=738). Also, in calculation of Rel_θ, we set ϑ=0.5. Due to space limit, we report the assessment of S2 against a consumer request γ=0.5 only as a representative case. Note that different QoS assessment methods exist and can be adopted to derive a QoS verdict (Ref), but which one to use will not affect how confidence is calculated in our work. For this reason, we assumed a simple QoS assessment in this paper: the average of all historical QoS data that are available. Figure 2 shows the result of this experiment.

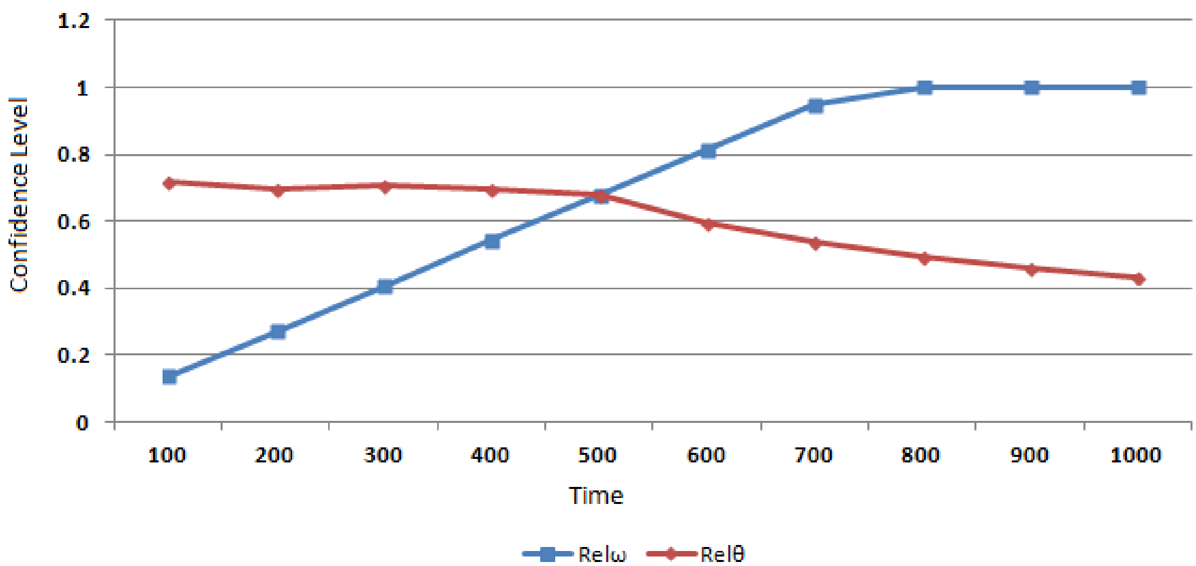

Figure 2: Assessment Confidence of S2

To observe the effect of different data sizes and deviations on assessment, we repeated our quality calculation for every addition of 100 service instances. In Figure 2, it is clear that the Rel_ω increased relatively with data size. That is, Rel_ω increased from 0 to 1 as the number of data items increased, and stayed at 1 when the data used in assessment exceeds m (i.e. 738).

On the other hand, Rel_θ can be described in terms of two phases. From t=0 to t=500, since S2 has delivered the required service level (i.e. 0.5) with less deviation, the Rel_θ was high (around 0.7). In the second phase (from t=500 to t=100), the Rel_θ has gradually decreased to 0.434 due to the fact that S2 has started delivering more fluctuated services (i.e. service deviation is equal to 0.20). This experiment illustrates that the inclusion of data size and deviation for confidence calculation has the desired effect.

Experiment 2 – Service Selection between multiple providers.

The goal of QoS assessment is to help a consumer to choose a service that offers the required quality. In this experiment, we demonstrate how our proposed confidence can improve service selection. We report service selection with and without our proposed confidence.

Figure 3 shows service selection without confidence. Since all four services delivered similar mean qualities (i.e. 0.5), the QoS assessment method made random selection of the services. Although S1 has deviated a lot in delivering the required level of service (i.e. 0.5), it has been chosen 30% of time. On the other hand, the service that delivered the required service consistently (S4) was selected only once (at 200).

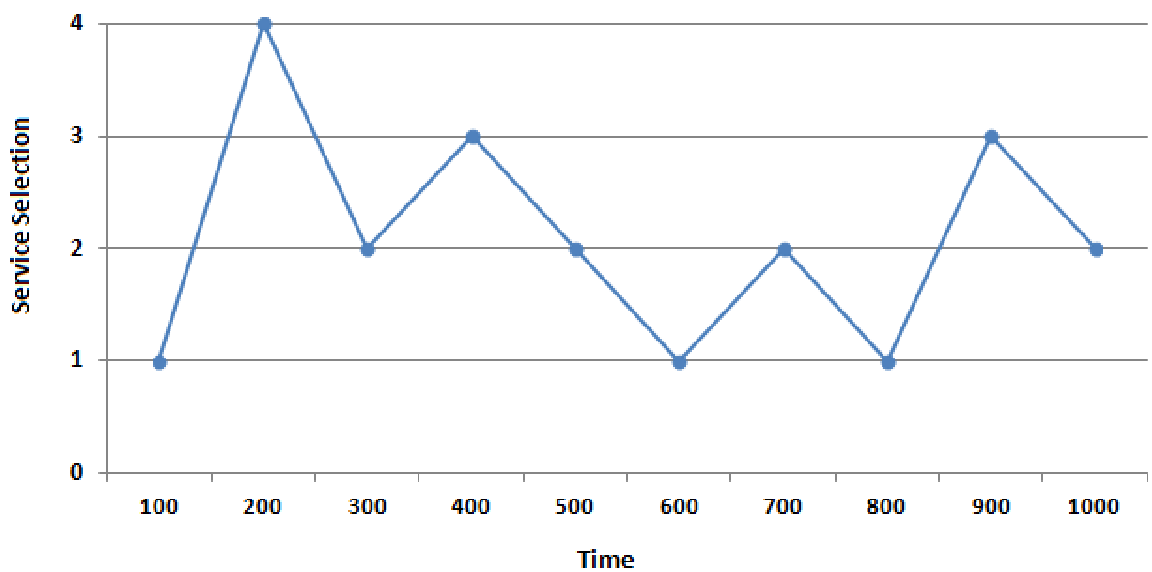

Figure 3: Service Selection without Confidence

Figure 4 shows service selection with confidence. The confidence levels of the four services are calculated using Equation 6 and plotted in Figure 5 for better explanation.

The behaviour observed in Figure 4 can be described in terms of two phases. From t=0 to t=500, S2, S3 and S4 were delivering similar service level (0.5) consistently, so the one with a higher confidence value was selected by the QoS assessment (see Figure 5). Note that, in contrast to the previous experiment, S1 was never selected because it had more fluctuated service delivery suggesting that the derived mean of S1 is a less reliable verdict. In the second phase (t=500 to t=1000), the assessment confidence of S2 and S3 slightly decreased due to more fluctuated service delivery and no more data for assessment, respectively. S4, on the other hand, kept delivering consistent service quality and, consequently nominated by the QoS assessment as the best service for the consumer. This experiment provides clear evidence that by including confidence as part of QoS assessment, a service consumer can be better guided in choosing the right service in a cloud computing environment.

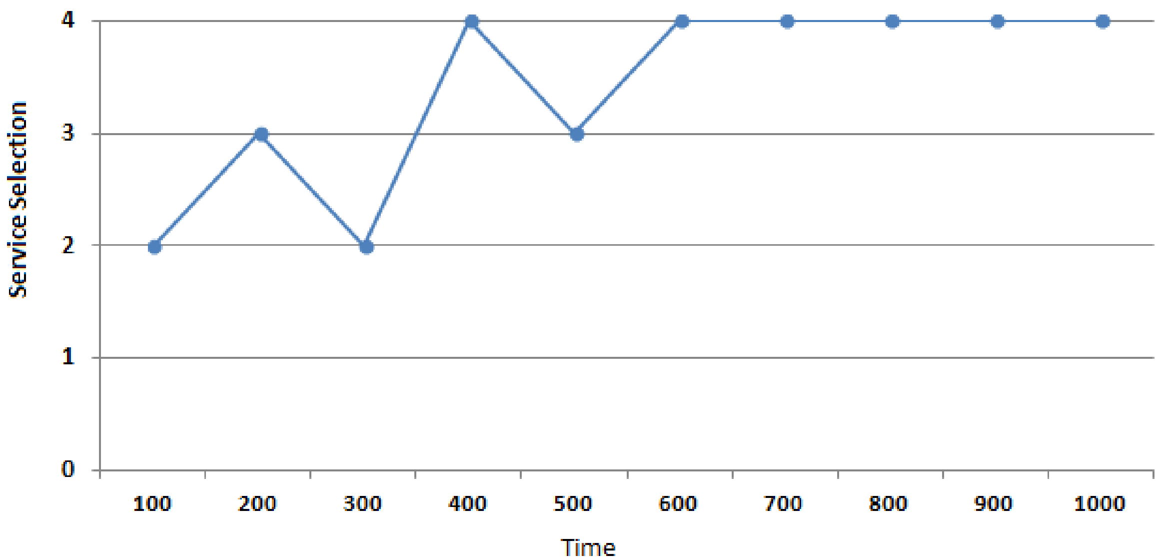

Figure 4: Service Selection with Confidence

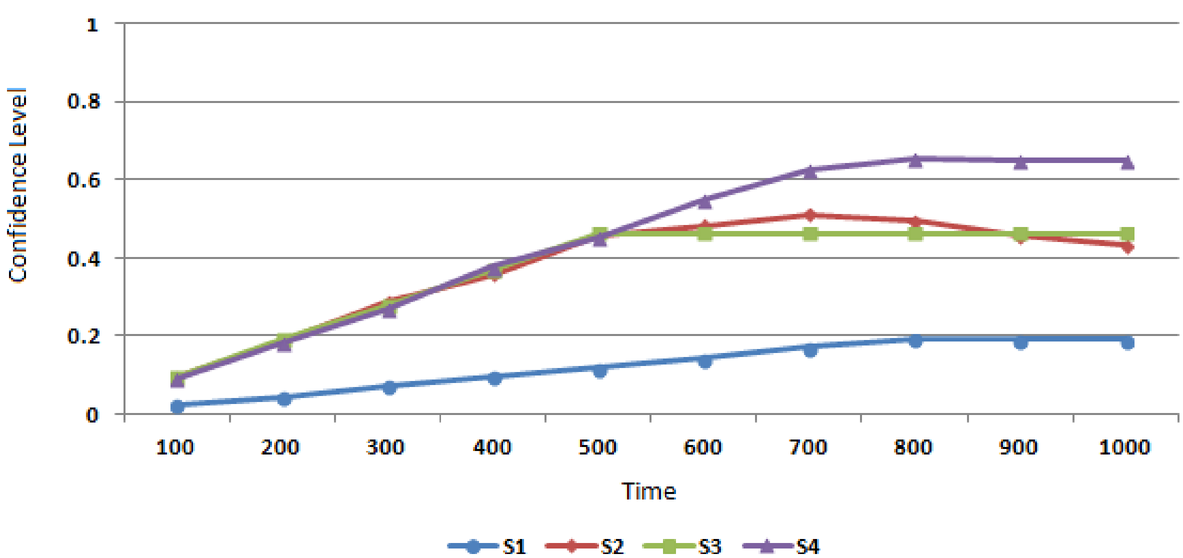

Figure 5: Confidence level of the four services ($\lambda = 0.95$ and $\varepsilon = 0.05$)

It is worth noting that our proposed confidence model is generic. That is, it is not limited to averaging all data as we did in our experiment, but can be integrated into other QoS assessment methods, for example, the methods proposed by (Liu. 2004, Shercliff. 2006, Al-Dossari. 2012b).

5. Conclusions

In this paper, we proposed a confidence model for QoS assessment that can be used to deal with uncertainty surrounding the reliability of a QoS prediction. The paper started by describing how decision theory may help address uncertainty that consumers may face in making a decision about which service to use. Then, to quantify confidence in QoS assessment, we presented a probabilistic model that integrates two reliability measures: the number of QoS data items used in assessment and the variation of data in the dataset. In this model, the data size measure indicates how strongly the prediction derived by the QoS assessment method is supported by the dataset or past evidence, and the data deviation measure indicates the service's consistency in delivering that prediction. Finally, we demonstrated through experiments that by adding confidence to the assessment process, consumers are able to make a better informed decision of which service to select, thus increasing their overall utility.

In future work, we intend to extend our proposed confidence model by using a "decaying" function, where more recent data will play a more important role in reliability measure. This is particularly important to assessing a service provider in terms of their behaviour change over time, for example, to gauge whether the quality of a service from a particular provider has been improving more recently.

Acknowledgement
The authors would like to thank King Saud University, Saudi Arabia, for their support in this work.

References
Al-Dossari, H., Shao, J., Preece, A. (2010) "QoS Assessment over Multiple Attributes". In Proceedings of the 10th International Conference on Computer and Information Technology, pp. 1456-1461.

Al-Dossari, H., Shao, J., Preece, A. (2012a) "Improving QoS Assessment over Multiple Attributes with Asynchronous Data". International Journal of Computer Systems Science and Engineering (IJCSSE 2012), CRL Publisher , Volume 27, pp. 133-147.

Al-Dossari, H. (2012b) "A Reputation-based Approach for Consumer Protection in Saudi Arabia". In 2012 Eighth International Conference on Next Generation Web Services Practices (NWeSP), pp. 321-326.

Deora, V., Shao, J., Gray, W.A, and Fiddian, N.J. (2003) "A quality of service management framework based on user expectations". 1st International Conference on Service-Oriented Computing (ICSOC 2003), Springer, pp. 104-114.

Huynh, T.D., Jennings, N.R., and Shadbolt, N.R. (2004) "FIRE: An integrated trust and reputation model for open multi-agent systems". 16th European Conference on Artificial Intelligence, IOS Press, volume 16, pp. 18–21.

Jøsang, A., and Ismai, R. (2002) "The beta reputation system. In 15th Bled Electronic Commerce Conference" Citeseer, volume 160.

Jureta, I.J., Herssens, C. and Faulkner. S. (2009) "A comprehensive quality model for service-oriented systems". Software Quality Journal, Springer, volume 17, pp. 65–98.

Liu, Y., Ng, A. H. H. and Zeng, L. (2004) "QoS Computation and Policing in Dynamic Web Service Selection". In 13th International World Wide Web Conference, pp 66-73.

Kim, D., Lee, S., Han, S., and Abraham, A. (2005) "Improving web services performance using priority allocation method". IEEE Computer Society, pp. 201–206.

Keung, S.N.L.C., and Griffiths, N. (2008) “Using recency and relevance to assess trust and reputation”. In Proceedings of AISB 2008 Symposium on Behaviour Regulation. The Society for the Study of Artificial Intelligence and Simulation of Behaviour.

Milton, J.S., and Arnold, J.C. (2002) “Introduction to probability and statistics: principles and applications for engineering and the computing sciences”. McGraw-Hill, Inc. New York, NY, USA.

Mui, L., Mohtashemi, M., and Halberstadt, A. (2002) “A computational model of trust and reputation”. In 35th Annual Hawaii International Conference on System Sciences (HICSS 2002), IEEE Computer Society, pp. 2431–2439.

Ristad, E.S. (1995) “A natural law of succession”. Arxiv preprint cmp-lg/9508012.

Sabater, J., and Sierra, C. (2001) “REGRET: reputation in gregarious societies”. In 5th International Conference on Autonomous Agents, ACM, pp. 194–195.

Sahai, A., Ouyang, J., Machiraju, V., and Wurster, K. (2001) “Specifying and guaranteeing quality of service for web services through real time measurement and adaptive control. E-Services”. Software Research Department, HP Laboratories, Palo-Alto, E-Service Management Project.

Shercliff, G., Shao, J., Gray, W. A., and Fiddian, N. J. (2006) ”A Multiple Quality-Space Mapping Approach to QoS Assessment”. In 6th International Conference on Computer and Information Technology, pp. 26-31.

Teacy, W.T.L., Patel, J., Jennings, N.R., and Luck, M. (2006) “Travos: Trust and reputation in the context of inaccurate information sources”. Autonomous Agents and Multi-Agent Systems, 12(2), pp. 183–198.

Svenson, O (1979) “Process descriptions of decision making”. Organizational behavior and human performance, 23(1), pp. 86–112.

Wohlstadter, E., Tai, S., Mikalsen, T., Rouvellou, I. and Devanbu, P. (2004) “Glueqos: Middleware to sweeten quality-of-service policy interactions”. IEEE Computer Society.

Zhang, J. and Cohen, R. (2006) “A personalized approach to address unfair ratings in multiagent reputation systems”. In 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2006) Workshop on Trust in Agent Societies.