1-1-2010

QoS probability distribution estimation for web services and service compositions

Huiyuan Zheng
Macquarie University

Jian Yang
Macquarie University

Weiliang Zhao
Macquarie University, wzhao@uow.edu.au

Recommended Citation

Zheng, Huiyuan; Yang, Jian; and Zhao, Weiliang: QoS probability distribution estimation for web services and service compositions 2010, 1-8.
https://ro.uow.edu.au/engpapers/5209

QoS Probability Distribution Estimation for Web Services and Service Compositions

Huiyuan Zheng
Department of Computing
Macquarie University
Sydney, Australia
huiyuan.zheng@mq.edu.au

Jian Yang
Department of Computing
Macquarie University
Sydney, Australia
jian.yang@mq.edu.au

Weiliang Zhao
Department of Computing
Macquarie University
Sydney, Australia
weiliang.zhao@mq.edu.au

***Abstract*—In this paper, we propose an approach to model the probability density function (PDF) of the QoS of a web service (QoWS) based on non-parametric statistical method. Mathematical formulas are designed to calculate the QoS distributions for service compositions (QoCS). Experiment has been done to show that the proposed QoWS distribution modeling approach is much more accurate than exiting methods. An accurate PDF estimation can be obtained for the QoCS if the PDF of the component QoWSs are modeled by the proposed method.**

***Keywords*-QoS; Web service; service composition; probability distribution;**

I. Introduction

The nature of services creates the opportunity for building composite services by combining existing elementary or complex services (referred to as component services) from different enterprises and in turn offering them as high-level services or processes (referred to as service compositions). QoS analysis becomes increasingly challenging and important when complex and mission critical applications are built upon services with different QoS. Thus solid model and method support for QoS predication in service composition become crucial and will lay a foundation in further analysis of complexity and reliability in developing service oriented distributed applications.

Describing QoWS accurately lays importance in the following aspects: (1) QoS needs to be reasonably specified in a service contract between a service provider and a consumer to guarantee the provider promised QoS; (2) The QoS of a service composition needs to be estimated based on the quality of individual web services to make sure that the composition can satisfy the expectations of end users.

Some QoWS metrics, such as response time, dynamically changes. Fixed value is not able to describe the QoWS effectively. For example, two web services having the same mean or maximum QoS value may have quite different QoS probability distributions. A service consumer may choose one web service over the other based on the QoS probability distributions, for example, the distribution with the smaller deviation would be preferred. A PDF is the best way to reflect the dynamic feature of a QoWS metric. A PDF of a QoWS is a function that describes the relative likelihood for this QoWS to occur at a given QoWS value.

PDF represented QoS has already been used in service level agreements (SLA) and service selections in service compositions [1,2], which shows promising selection result over traditional fixed QoS value methods. Standard statistical distributions such as T Location-scale and Normal distributions are applied to represent the PDF distributions for QoWSs. But as standard statistical distributions have regular shapes, they can only model the main body of the QoS distribution accurately. As is shown in the following example, other part of a distribution though taking a small proportion can generate a significant effect on the QoS distribution of a service composition and therefore cannot be ignored.

A. An Example

We will use the QoS metric response time as an example to illustrate the necessity of accurate estimation of QoWS. Figure 1 is the PDF distributions of the response time of four web services WS1, WS2, WS3, and WS4. We will first talk about the circumstance of one web service being invoked in a service composition for multiple times. Then, we will talk about the circumstance of multiple web services being invoked in a service composition.

Figure 2 shows the response time distributions for WS1 being sequentially invoked from 2 to 5 times in a service composition. It can be seen from the response time distribution of WS1 (see Figure 1) that most observations of response time are ranging between 0.5×10^4 to 1×10^4 ms (referred to as `main part`). There are some sparse observations over the smaller (less than 0.5×10^4 ms) and larger (more than to 1×10^4 ms) response time (referred to as `other part`). These sparse observations only take quite a small proportion of the whole response time observations. When WS1 is invoked in a service composition, the more times it is invoked, the smaller proportion the main part of the PDF takes in the response time distribution of the service composition (see Figure 2). It means that the effect of the main part of the QoS distribution of WS1 on the QoS distribution of the service composition decreases. It can be seen from the top rightmost plot of Figure 2 that the

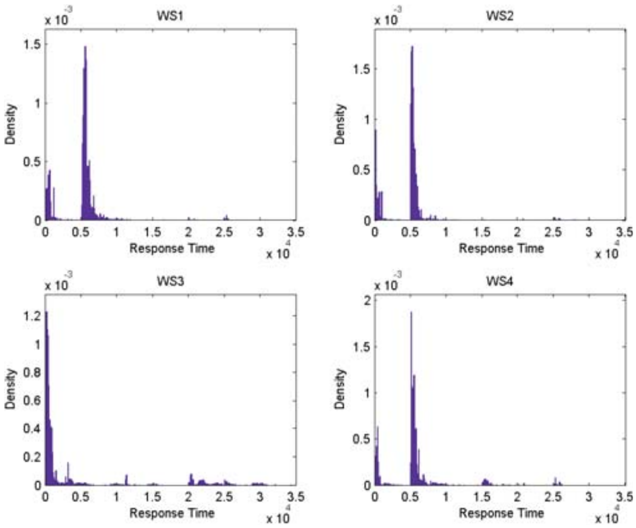


Figure 1. PDFs of Four Web Services

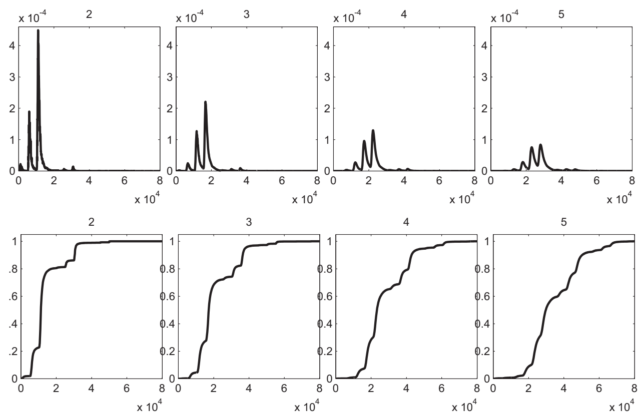


Figure 2. Composition of WS1 2 to 5 Times

response time distribution becomes bimodal (the maximum probability density is attained at two response time values) when WS1 is invoked 5 times.

Figure 3 illustrates the response time distribution for web services WS1, WS2, WS3, and WS4 being sequentially invoked in a service composition. From Figure 3, it can be seen that the distribution converges to zero density slowly. There is a heavy tail from 2×10^4 to 6×10^4 ms and there are three bumps in the tail. This tail takes a considerable proportion of the whole response time observations and can not be ignored.

Currently, QoWS are modeled as (1) single constant values [3,4], such as the mean, the minimum, or the maximum QoS value; (2) probability mass function (PMF) [5]. For example, $P(cost = 100dollars) = 0.4$, $P(cost = 200dollars) = 0.6$ is a PMF for the cost of a web service; and (3) standard statistical distributions [2,3], such as *normal distribution* and *t location-scale distribution*. From the example shown through Figure 1, 2, and 3, the limitations of exiting QoS modeling methods can be summarized as follows:

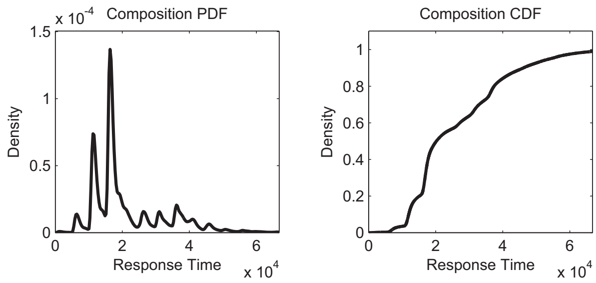


Figure 3. Composition of WS1 to WS4

•As the QoS of a web service changes dynamically, single constant value-modeled QoS is not able to reflect this variability;

•The values of some QoS metrics, such as response time, range continuously. PMF-modeled QoS is not able to model this characteristic;

•Standard statistical distribution modeling method can only model distributions with regular shapes and smooth tails. It is not able to model the minor part (i.e. the part takes quite a small proportion in a distribution) of a distribution precisely (proof can be seen in the experiment in Section V). If the QoWS distributions of component web services are not accurately modeled, the estimated QoS for the service composition based on these component QoWSs would be quite misleading. In addition, from the example discussed before, we have already known that the minor part of a QoS probability distribution cannot be ignored because it will have a significant effect on the QoS probability distribution of a service composition when one web service is kept being invoked in a service composition (for example in a loop pattern) or multiple web services are invoked together. Therefore, standard statistical distributions are not suitable for QoWS modeling.

B. Contributions of This Paper

In this paper, we use a non-parametric statistical method to estimate the PDF for QoWS. This method is distribution free, which do not rely on assumptions that the QoS data are drawn from a given probability distribution. This property makes this method more robust and accurate than the ones used in existing work [2,6] on QoWS estimation.

A general QoS calculation approach for service compositions has been explored in this work. No matter the QoSs of component web services are modeled into single values, PMFs, or PDFs, the proposed method is able to calculate the QoS for a service composition.

Simulation has been done to show that (1) the proposed PDF modeling method can fit to the QoWS sample very well and is more accurate than existing modeling methods;

(2) when PDFs of QoWSs are generated by the proposed non-parametric method, the calculated PDF of QoCS can fit to the simulated QoCS sample accurately.

The remainder of the paper is organized as follows: Section II discusses the related work. In Section III, QoWS modeling approaches are introduced in detail. Section IV discusses a general QoS calculation method for a service composition. Experiments are done in Section V to compare different QoWS modeling approaches. Section VI concludes the work.

II. Related Work

A. QoS Modeling for Web Services

In [4], QoS metrics, such as execution duration and reputation, are estimated as the mean value of the past observation. QoS metrics, such as successful execution rate and availability, are estimated statistically. For example, the transmission time is estimated as the mean value of the past observation of the transmission time, while the success rate is estimated as the number of times that a service has been successfully completed divided by the total number of invocations.

In [3], the estimation of QoS metrics are divided into two classes: basic and distributional. The basic estimation of QoS metrics corresponds to the minimum, average, and maximum QoS values associated with the execution of a task. The distributional estimation corresponds to the specification of a constant or of a standard statistical distribution function (such as exponential, normal, Weibull, and uniform) which statistically describes a task behavior at runtime. When a distribution function is unpractical to be derived, a histogram is recommended against an analytical formula.

In [5], QoS metrics, such as cost and reliability, may have a set of QoS values and can be modeled as PMFs naturally. For example, the execution of a web service may cost differently which is: 100 dollars, 200 dollars, or 300 dollars. When the probability of each cost value is estimated, the PMF for the cost of the web service is generated. For QoS metrics, such as response time, the range of a QoS metric can be divided into discrete subintervals. Then after the probability for each subinterval is estimated, the PMF for the QoS metric can be obtained.

In [2], the response time of web services are measured and used as sample data. Parametric statistics is used to make inferences about the parameters of the distribution. Specifically, the distribution of a web service is assumed to be a T Location-scale distribution. Based on the sample data of each web service, the parameters of the T Location-scale distribution are estimated. Then this distribution is treated as the PDF of that web service.

In [1], a probabilistic approach is proposed to select web services whose QoSs are modeled as probability distributions. The response time of a web service that is under a certain load is supposed to follow a Normal distribution. The response time of a web service is dependent on the load which varies throughout the week. When the Normal distributions of a web service under different load are known, the distribution of the response time of the web service is the aggregation of these Normal distributions which results in a probability distribution following no obvious pattern.

B. QoS Calculation for Service Compositions

For single constant value represented QoS, aggregation method [3,7] is proposed to calculate the composite QoS. A composition can be regarded as being composed of composition patterns. Formulas to calculate QoS for these patterns are given. But these formulas can only be applied to single values.

For QoS represented by PMF, the calculation method is much the same as it is for single values [5]. The difference is that the probability of each possible QoS value of the composite service needs to be taken into account.

For standard statistical distribution represented QoS, simulation approaches are applied to compute the composite QoS [2,8].

Mathematical formulas are developed in [9] to calculate the throughput for composition patterns with execution time represented by distributions.

[10] presents a tool for predicting composite QoS. Component QoS can be modeled as single value or parameters of standard statistical distribution. But this tool does not support complex patterns such as loop.

III. Approach

In this section, we will first study the QoWS distribution generation approach adopted by existing work. Then, we will introduce the QoWS distribution generation method to be used in this paper.

A. Distribution Generation Based on Parametric Statistical Approach

One QoS value of a QoS metric can be obtained after per execution of a web service. All these QoS values of per QoS metric are stored in a log file of the web service and can be used as a *sample* to generate a probability distribution for the specific QoS metric of the web service.

Methods adopted in existing work [2,6] assume that a *sample* comes from a type of a standard statistical distribution which can be represented by several parameters. For example, a Normal distribution can be represented by two parameters which are the mean (μ) and the variance (σ^2) of the sample [11]. To get the QoS probability distribution, they make inferences about the parameters of the distribution. Since every probability distribution in this kind of methods is assumed to be represented by a set of parameters, we refer this kind of methods as *parametric methods*. The advantage of parametric methods is that the parametric formulae are often simple to write down and fast to compute. However, parametric methods are not robust, because they make more assumptions and if those assumptions are incorrect, parametric methods can be very misleading [11].

Next we will study how to obtain the QoS probability distribution from the QoS *sample* based on parametric method [12].

1) Assumption of the type of the probability distribution for the sample: The *sample* is assumed to follow a certain distribution (such as Normal distribution), which is called *hypothesis* H_0. In general, the QoS distribution should be in the form that it has a long tail and the very large values have small frequency while the intermediate values have large frequency. Distributions, such as log-logistic, gamma, t location-scale, etc., have this characteristic.

2) Estimation of the parameters for a probability distribution: Given the sample and the *hypothesis* H_0, there are many estimation methods can be adopted to obtain the parameters of the distribution. In this paper, *maximum likelihood estimation* (MLE) [12] method is adopted. The discussion of the estimation theory is out of the scope of this paper.

When the parameters of the hypothesized distribution are computed, the QoS distribution of a web service is obtained.

B. Distribution Generation Based on Non-Parametric Statistical Approach

We will use Gaussian Kernel Density estimation in this paper to generate QoWS probability distribution. As Gaussian Kernel Density estimation is a non-parametric way of estimating the PDF of a random variable, we refer this method as *non-parametric method* in the rest of the paper.

In this method, if $x_1, x_2, ..., x_n$ is an independent and identically-distributed sample of a QoS metric of a web service, then the approximation of the PDF of the sample is [13]:

$$\hat{f}(x) = \frac{1}{nh} \sum_{1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}} \tag{1}$$

where h is a smoothing parameter and can be calculated as follows:

$$h = 1.06\sigma n^{-1/5} \tag{2}$$

where n is the size of the sample and σ is the standard deviation of the sample.

C. Test the goodness of fit between the estimated QoS distribution and the sample

In this paper, Chi-square Test is applied to compare the difference between the estimated QoS distribution and the sample. The formula of Chi-square test is as follows:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} \tag{3}$$

where n is the size of the sample, k is the number of disjoint intervals that the sample is divided into, f_i is the *observed frequency* that represents the number of sample that is within the interval i (i=1,2,...,k), and p_i is the probability within the interval i according to the estimated PDF.

The value of χ^2 represents the discrepancy between the *sample* and the estimated PDF of QoWS. In the experiment of this paper (Section V), χ^2 will be used as a criterion to measure the performance of the QoS distribution estimation methods.

IV. A General QoS Calculation Approach for a Service Composition

A composite service is regarded as being constructed based on four composition patterns, i.e. sequential, parallel, conditional, and loop. The formal definitions and modeling methods of these patterns and the regressive processing method of a service process based on these basic patterns have been discussed in [14].

In this section, we design a calculation approach which can compute the QoS probability distributions for service compositions. In this approach, we assume that the QoSs of web services are independent of each other. The QoS metric response time is taken as an example here. The QoS calculation formulas for different composition patterns are listed as follows:

A. Sequential Pattern

The response time of a Sequential Pattern is the sum of the QoS of its component web services. The QoS of a Sequential Pattern is the convolution of the PDFs of the component QoWSs,i.e.

$$f(q) = (f_1 * f_2)(q) = \int_0^q f_1(x) f_2(q-x) dx \tag{4}$$

where $f(q)$ is the PDF of response time of a Sequential Pattern, $f_1(q)$ and $f_2(q)$ are the PDFs of the component QoWSs.

B. Parallel Pattern

The response time of a Parallel Pattern with synchronized merge is the maximum response time of its component web services. The probability distribution can be calculated as follows:

$$F(q) = \prod_{i=1}^{n} F_i(q) \tag{5}$$

$$f_(q) = \sum_{i=1}^{n} f_i(q) \prod_{j=1,\ldots,n \& j \neq i} F_j(q) \tag{6}$$

where $f(q)$ and $F(q)$ are the PDF and Cumulative Distribution Function (CDF) of the response time of a Parallel Pattern; $f_i(q)$ and $F_i(q)$ are the PDF and CDF of the response time of component service i; and n is the number of component services within this pattern.

C. Conditional Pattern

The response time of a Conditional Pattern is the probability weighted sum of the response time of its component web services. The QoS distribution of a Conditional Pattern can be calculated as follows:

$$f(q) = \sum_{i=1}^{n} p_i f_i(q) \tag{7}$$

where $f(q)$ is the PDF of the response time of a Conditional Pattern; n is the number of component services within this pattern; $f_i(q)$ is the PDF of the QoS of component service i; and p_i is the execution probability for component service i.

D. Loop Pattern

In [7], we have given detailed discussion on the structure analysis method for an arbitrary Loop Pattern to compute its QoS. To sum up the method in [7], statistically, a Loop Pattern can be seen as a Conditional Pattern with a Sequential Pattern in each path. With calculation formulas for the execution probability of each path of the Conditional Pattern given in [7] and the formulas of computing the response time probability distribution of a Sequential Pattern and a Conditional Pattern known (Formulas 4 and 7), the distribution of the response time of a Loop Pattern can be obtained.

E. Computational Complexity of the Proposed QoS Calculation Approach

The computation of convolution takes most time in calculating the QoS of a service composition. If convolution is computed directly, it is often too slow to be practical. In this paper, with the help of the convolution theorem [15] and the fast Fourier transform (FFT) [16], the complexity of the convolution is reduced from $O(n^2)$ to $O(nlogn)$ [15]. A convolution of f_1 and f_2 is calculated as follows:

$$f_1 * f_2 = F^{-1}\{F\{f_1\} \cdot F\{f_2\}\} \tag{8}$$

where F represents Fourier transform while F^{-1} represents inverse Fourier transform which can be performed by FFT algorithm efficiently.

After the analysis of the complexity for convolution computation, now we can get the computational complexity for the proposed QoS calculation approach which is $O(mnlogn)$ (m is the number of web services in a service composition and n is the number of discrete points in a QoS distribution).

V. EXPERIMENT

A. QoS Distribution Generation for Web Services

In this subsection, we compare using parametric and non-parametric approaches to estimate the QoS (response time) distributions for two web services: *Random Image* and *Dilbert*. The QoS samples of the two web services are from WS-DREAM dataset [17].

Table I
CHARACTERISTIC RESPONSE TIME (MS)

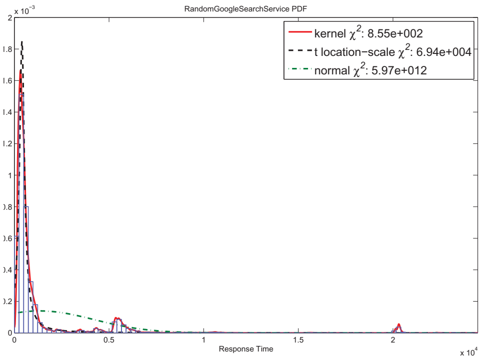
		mean	CDF=90%	CDF=95%	CDF=99%
Random	S	1339.6	4345	5666	20255
	K	1339.0	4394	5691	20255
	T	645.8	24489	24489	24489
	N	1935.5	24244	24244	24244
Dilbert	S	738.29	941	2855	5570
	K	737.90	941	3084	5603
	T	511.85	1284	16367	16367
	N	914.29	9840	9840	9840

S: Sample. K: Non-parametric method. T: T location-scale distribution
N: Normal distribution

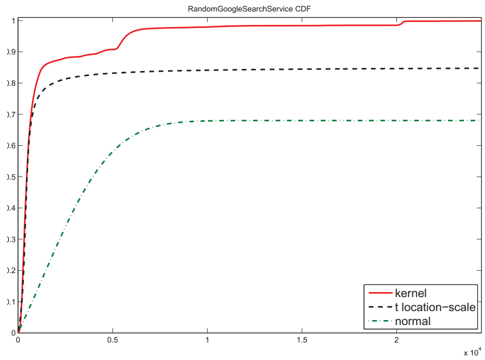
In Figure 4, histograms in Figure 4(a) and 4(c) represent the QoS samples of web services *Random Image* and *Dilbert* respectively, solid curves represent QoS distributions (PDF or CDF) generated by non-parametric approach, dashed curves represent T Location-scale distributions of the QoS generated by parametric approach, and dash-dotted curves represent Normal distributions of the QoS generated by parametric approach. It can be seen that the solid curves fit the QoS sample very well both in the main part and in the tail part. Small bumps in the tail part are detected by the non-parametric approach. The T Location-scale distributions can only fit the main part of the QoS sample well. As for the Normal distributions, their shapes are far from reflecting the real distributions. Most importantly, it can be seen from Figure 4(b) and 4(d) that the CDFs of T Location-scale and Normal distribution are not 1 when response time approaches infinity. This is because the variance of the sample is so large that the left part of the PDF of Normal and T Location-scale distributions goes to negative x-axis which leads to the integration of the PDF over the positive x-axis is less than 1. Therefore, it can be foreseen that the QoS estimation for a service composition would be quite misleading when the QoSs of its component services are generated from parametric method. This will be further proved by experiment in Section V-B.

So far, we have analyzed the distributions in Figure 4 visually. Next, we will analyze them quantitatively. First of all, the discrepancies χ^2 between the estimated distributions and the sample have been indicated in Figure 4(a) and 4(c). It can be seen that the distributions estimated by non-parametric approach have the smallest discrepancies with the samples. The discrepancies generated by T Location-scale and Normal distributions are quite large. Next, we evaluate response time at four characteristic points (referred to as *characteristic response time* in the following part) which are the mean response time and the response time at three CDF percentiles, i.e. 90%, 95%, and 99%. The quantitative results are shown in Table I.

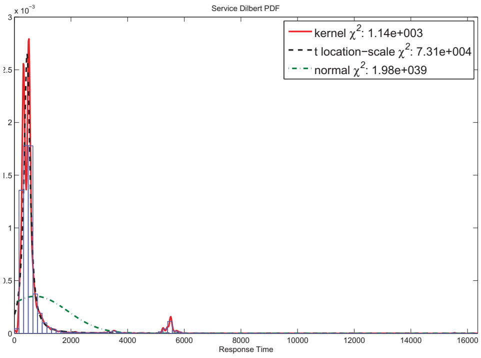
It can be seen that the QoS distribution generated by non-



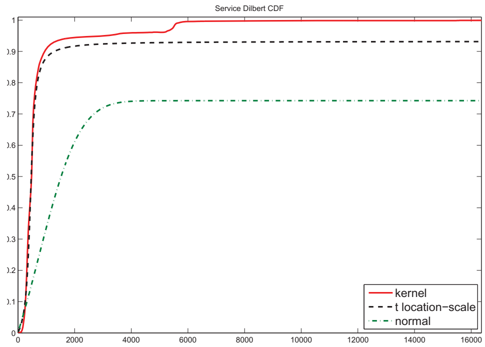
(a) Random Image PDF



(b) Random Image CDF



(c) Dilbert PDF



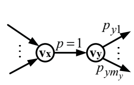
(d) Dilbert CDF

Figure 4. QoS Distribution Generation

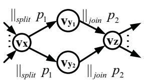
parametric approach has the closest characteristic response time as that of the sample. As to the T Location-scale and Normal distributions, they are not even able to give the approximate characteristic response time values. For example, if the user requires that for 95% of chance the response time of a web service should be less than 6000 ms. As is seen from Table I, web service *Random Image* and *Dilbert* have a response time of less than $5666ms$ and $2855ms$ respectively (seen from the sample) for 95% of chance. In fact, both of the two services meet the user requirements on response time. This characteristic can be reflected accurately by the QoS distribution obtained by non-parametric approach. However, in parametric method generated T Location-scale and Normal distributions, for 95% of chance the response time are all far more than 6000 ms for both web services. If these distributions were used in SLA, it would lead to quite pessimistic contract and cause the loss of the service provider. This further proves that parametric approach is not suitable for QoS probability generation for web services.

One thing to be noted is that the mean response time getting from Normal distributions is not equal to the mean response time of the QoS samples as expected. This is because response time is all positive values and only the positive part of the Normal distribution is taken into account when the mean response time is calculated. When the variance of the sample is large enough to make the Normal distribution expand to the negative x-axis, the mean value will of course shift rightwards, i.e. larger than it is supposed to be.

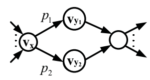
To sum up, QoWS distributions obtained by non-parametric approach are able to represent the real QoWS distributions while standard statistical distributions do not have this ability.



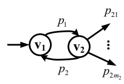
(a) Sequential



(b) Parallel



(c) Conditional



(d) Loop

Figure 5. Four Composition Patterns

B. QoS Distribution Calculation for Different Patterns

In this subsection, we will show the accuracy of the QoWS distributions obtained by non-parametric approach from a service composition point of view. We use the

Table II
MEAN RESPONSE TIME OF DIFFERENT APPROACHES

	Sequential	Parallel	Conditional	Loop
S	2054.7	1681.0	1085.4	6997.7
M	2077.8	1339.6	1099.0	6926.1
K	2054.9	1688.1	1098.5	6736.5
T	1253.7	970.2	682.7	2248.1

S: Simulation; M: Mean; K: Non-parametric method; T: T Location-scale

two web services in Figure 4 as component web services of a Sequential Pattern, a Parallel Pattern, a Conditional Pattern, and a Loop Pattern respectively (see Figure 5). Monte Carlo simulation method is used to simulate the QoSs of the four composition patterns. For each time of simulation, the response time of the two web services are generated randomly from their QoS samples (see detailed about these samples in Section V-A). The QoS for the Sequential, Parallel, Conditional, and Loop Pattern is the sum of response time, the maximum of response time, one of the response time depending on which path is taken, and the total aggregated response time depending on the times that the loop takes in that particular simulation respectively. The simulation is executed for 20,000 times and therefore 20,000 response time for each of the composition pattern is obtained. The simulation results are shown as histograms in Figure 6. Solid curves represent the calculated QoS distributions of composition patterns when the component QoWSs are non-parametric approach generated distributions. Dashed curves represent the calculated QoS distributions of composition patterns when the component QoWSs are T Location-scale distributions. As Normal distributions are far from accurate in representing QoWSs, we did not consider Normal distributions here.

It can be seen from Figure 6 that the calculated QoS distributions based on non-parametric approach generated QoWS distributions represent the simulation results quite well not only for the main part but also for the tail part of the distributions. The calculated QoS distributions based on T Location-scale QoWS distributions are not able to represent the tale part accurately.

The quantitative results are shown in Table II and III for different composition patterns.

Table II is the mean response time for Sequential, Parallel, Conditional, and Loop Patterns. It can be seen that the mean response time of the four patterns calculated based on both mean value modeled QoWSs and non-parametric approach generated QoWSs are quite close to the mean values of the simulation results. The mean values of the calculation results based on T Location-scale modeled QoWSs are quite different from those of the simulation results.

Table III shows the response time of different patterns at $CDF = 90\%$, $CDF = 95\%$, and $CDF = 99\%$ respectively. It can be seen the calculated response time at those characteristic points are quite close to the simulated response time when the QoS distributions of component web services are generated by non-parametric approach.

According to the above experimental results and analysis, we can conclude that non-parametric approach modeled QoWS distribution represents the QoS distribution of a web service accurately and when it is used in representing QoS of component web services, the calculated QoS distribution of a service composition is trustable.

VI. CONCLUSION

In this paper, we introduce two statistical methods of estimating the QoS distributions for web services: parametric statistical method which has been used in some work and non-parametric statistical method which is first proposed to be used in QoWS distribution estimation in this paper. Experimental results show that the non-parametric approach can estimate QoWS distribution much more accurately and is able to reflect the real QoWS distribution.

More and more services (i.e. service composition) are developed upon the composition of existing web services. It is meaningful to estimate the QoS of a service composition at design time. We give a set of formulas for the calculation of the QoS distribution of a service composition with its component QoWSs modeled as PDFs. Simulation results show that the QoS of a service composition can be accurately estimated when its component QoWSs are represented by the PDFs getting from the non-parametric statistical method.

REFERENCES

[1] A. Klein, F. Ishikawa, and B. Bauer, "A probabilistic approach to service selection with conditional contracts and usage patterns," in *ICSOC-ServiceWave '09*, 2009, pp. 253–268.

[2] S. Rosario, A. Benveniste, S. Haar, and C. Jard, "Probabilistic qos and soft contracts for transaction-based web services orchestrations," *IEEE Trans. Serv. Comput.*, vol. 1, no. 4, pp. 187–200, 2008.

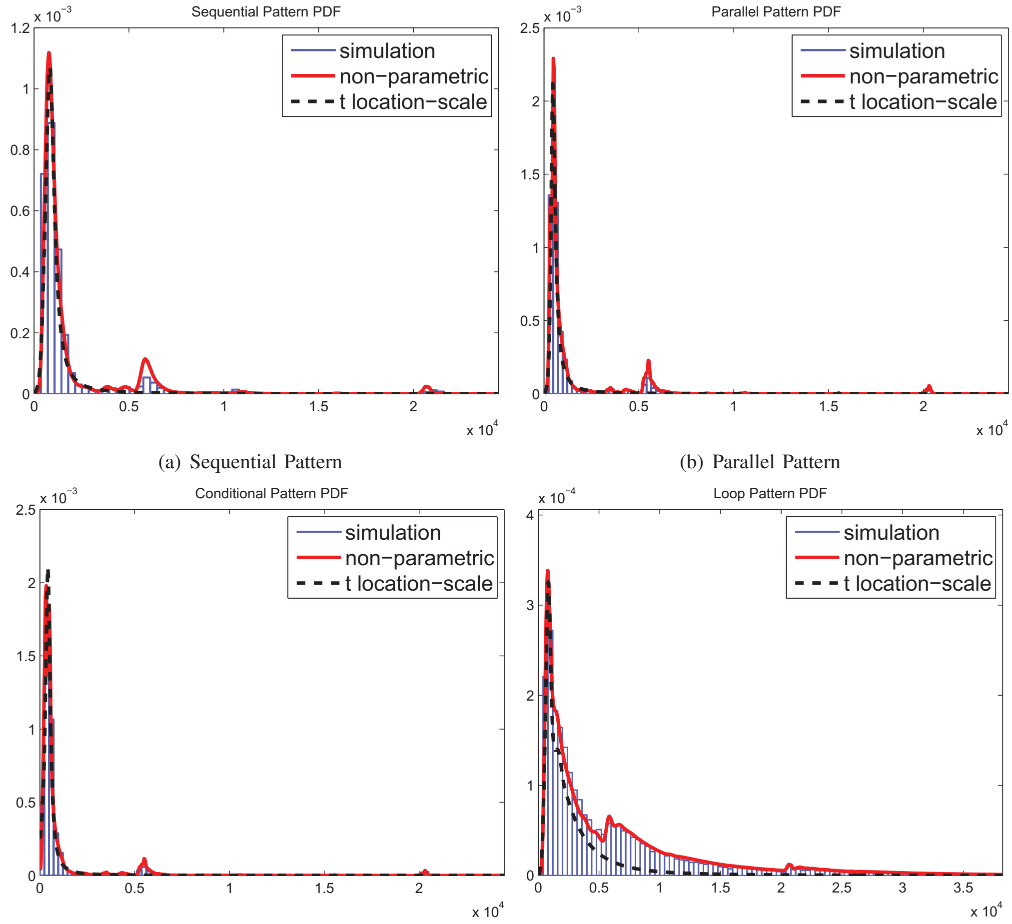
[3] J. Cardoso, J. Miller, A. Sheth, and J. Arnold, "Quality of service for workflows and web service processes," *Journal of Web Semantics*, vol. 1, pp. 281–308, 2004.

[4] L. Zeng, B. Benatallah, A. Ngu, M. Dumas, J. Kalagnanam, and H. Chang, "Qos-aware middleware for web services composition," *IEEE Transactions on Software Engineering*, vol. 30, no. 5, pp. 311–327, May 2004.

[5] S.-Y. Hwang, H. Wang, J. Tang, and J. Srivastava, "A probabilistic approach to modeling and estimating the qos of web-services-based workflows," *Inf. Sci.*, vol. 177, no. 23, pp. 5484–5503, 2007.

[6] W. Wiesemann, R. Hochreiter, and D. Kuhn, "A stochastic programming approach for qos-aware service composition," in *CCGRID*, 2008, pp. 226–233.

[7] H. Zheng, W. Zhao, J. Yang, and A. Bouguettaya, "Qos analysis for web service composition," in *SCC*, 2009, pp. 235–242.



(a) Sequential Pattern

(b) Parallel Pattern

(c) Conditional Pattern

(d) Loop Pattern

Figure 6. QoS Distributions for Different Composition Patterns

Table III
CHARACTERISTIC RESPONSE TIME OF DIFFERENT APPROACHES

	Sequential		Parallel		Conditional		Loop	
	Simulation	Non-para	Simulation	Non-para	Simulation	Non-para	Simulation	Non-para
CDF=90%	5820	5820	5420	5380	1740	1860	17235	16940
CDF=95%	6420	6420	5780	5740	5460	5460	24058	23580
CDF=99%	20580	20620	20180	20180	10860	11340	38945	36870

[8] G. Grabarnik, H. Ludwig, and L. Shwartz, "Management of service process qos in a service provider - service supplier environment," in *CEC/EEE*, 2007, pp. 543–550.

[9] G. Grabarnik, H. Ludwig, and L. Shwartz, "Dynamic management of outsourced service processes' qos in a service provider - service supplier environment," in *BDIM*, 2008, pp. 81–88.

[10] C. Hughes and J. Hillman, "Qos explorer: A tool for exploring qos in composed services," in *ICWS*, 2006, pp. 797–806.

[11] http://en.wikipedia.org/wiki/Parametric_statistics#cite_note-Geisser_and_Johnson-0.

[12] M. Spiegel, A. Srinivasan, and J. Schiller, *Schaum's outline of theory and problems of probability and statistics*, 2nd ed. New York: McGraw-Hill, 2000.

[13] B. Silverman, *Density Estimation for Statistics and Data Analysis*, 1st ed. Chapman and Hall, 1986.

[14] H. Zheng, J. Yang, and W. Zhao, "Qos analysis and service selection for composite services," in *SCC*, 2010, pp. 122–129.

[15] http://en.wikipedia.org/wiki/Convolution_theorem.

[16] http://en.wikipedia.org/wiki/Fast_Fourier_transform.

[17] Z. Zheng, H. Ma, M. R. Lyu, and I. King, "Wsrec: A collaborative filtering based web service recommender system," in *ICWS*, 2009, pp. 437–444.