WSrep: A Novel Reputation Model for Web Services Selection*

Zhen Li, Sen Su, and Fangchun Yang

State Key Lab. of Networking and Switching, Beijing University of Posts and Telecommunications
187#, 10 Xi Tu Cheng Rd., Beijing, 100876, P.R. China
jasonlizhen@gmail.com, {susen, fcyang}@bupt.edu.cn

Abstract. Web services selection is based on QoS and trust. As one of the important attributes of QoS, reputation is commonly used to assess the trustworthiness of the web services and minimize the threats of transactions. However, most existing reputation models of web services are all based on the subjective user ratings. These systems are easily attacked by malicious raters. This paper presents a novel reputation model named WSrep, in WSrep, the reputation integrates user ratings and a significant objective factor-credibility of QoS advertisements which is an objective view of the past behaviors of a given service. Other contributions of the paper include a customer measurable QoS model, a Bayesian learning model for building the credibility, and a set of experiments to show the benefits of our approach.

1 Introduction

Web service technologies promise the dynamic construction of loosely coupled information systems [4]. As a consequence of the rapid growth of web services (especially functional overlapping services), systematically and (semi-) automatically selecting the "best" service becomes a difficult and challenging work. In such a scenario, the quality of service (QoS) and trust are the key factors to do services selection processes [7]. QoS can help customers to select a distinguished service that has higher qualities, and trust is used to assist customers to choose good providers who always deliver promised qualities honestly. As an important attribute of QoS, reputation is a measure of the trustworthiness of the services, it mainly depends on end user's experiences of using the service [8]. Reputation not only can be efficiently used to find good services, but also can stimulate transactions to be executed exactly without the expense of third party monitoring.

Recognizing the importance of reputation, an immediate question to ask is how to model the reputation. There is an extensive amount of research focused on building reputation models for web services [5], [6], and [9]. Further, some well-known sites

* This work is supported by the National Basic Research and Development Program (973 program) of China under Grant No. 2003CB314806; the program for Changjiang Scholars and Innovative Research Team in University (PCSIRT).

N.T. Nguyen et al. (Eds.): KES-AMSTA 2007, LNAI 4496, pp. 199–208, 2007.

such as bindingpoint.com [2] also provide an interface to rate web services. These approaches all defined reputation as average of subjective user ratings. Actually, these rating-based reputation systems do not work efficiently under some complex circumstances. A key limitation of current approaches is that they can not cope with the malicious rating attack [3]. In the web services settings, the dishonest providers often oscillate their reputation between building and milking to mislead the customers, they also collude with malicious raters who always provide low feedbacks to their opponents and high feedbacks to themselves. So the current subjective reputation models are hard to reveal the correct performance history of a given service. In addition, current approaches are not flexible enough to express a given customer's intentions and norms. They disseminate the values of services' reputation among different customers. These values represent the general perspectives of raters but not a personal opinion, it may make a customer to select a service with which he is not satisfied although this service may have a higher reputation value. So there is the need to model reputation from a personal perspective.

With these research problems in mind, we develop WSrep, a novel reputation model for web services selection. In WSrep, the reputation model is made up of the average of user ratings and the objective view of performance history of a given service. We propose a concept named "credibility of QoS advertisements" (Cre) as the objective factor. Cre is a probability which is used to predict the possibility of providers delivering QoS values complying with their advertisements. The Cre can reduce the wicked influence made by malicious rating attack and our reputation is computed at user local side with personal information for expressing the different cognitions of a given service from different customers. We also introduce a customer measurable QoS model for software agents to justify the compliance of the actual QoS values and their advertisements automatically, a Bayesian learning approach to model the Cre. Finally, we demonstrate the effectiveness of this approach via simulation-based experiments.

The remainder of this paper is organized as follows: Section 2 introduces our customer measurable QoS model. Section 3 discusses the model of WSrep. Section 4 presents a series of simulation-based experiments to show the effectiveness and benefits of our approach; Section 5 summarizes the related works in this field; Finally, Section 6 concludes the paper.

2 A Customer Measurable QoS Model

Many literatures like [11] and [12] introduce their comprehensive QoS attributes models which include service level attributes and network level attributes. However, a given customer can not validate the actual values of some attributes directly (e.g., throughput, which is the total number of served requests in a time window. etc.). In this section, we propose a customer Measurable QoS model-MQ used by customers' agents to justify the compliance of the actual QoS values and their advertisements automatically.

Definition 1: MQ

$$MQ \equiv \{Latency, DomQ\}$$

- *Latency:* It is the delay between sending a request and receiving a response.
- *Domain-specific QoS ($DomQ$):* Domain-specific QoS is a set of properties which is used to describe the characteristics of services in a special domain. It should be introduced by domain experts, for example, the "interest rate" is one of the domain-specific attributes of a loan service.

Latency can be justified by run-time monitoring; the values of domain-specific QoS are offered in the service results, user agents can parse these XML-based messages to get the real values and compare them with advertised ones automatically.

3 The WSrep Model

Modeling and designing WSrep are the main purposes of this paper. The WSrep mixes the subjective and objective view of past behaviors of providers, and uses this knowledge to help customers to select the most trustworthy service.

3.1 WSrep Parameters

In WSrep, We identify two important parameters:

- *User Ratings.* After a customer accessing a web service, he rates the service depending on his own preference. The ratings imply the level of satisfaction of users about the whole QoS attributes the provider delivered. So the user ratings are valuable for any reputation models (there are no other ways to express customers' satisfaction properly). User agents in WSrep can adopt any of the existing non-negative integer rating mechanisms (i.e. the rating mechanism of [2], where customers can rate a service with one integer from 0 to 10). This design makes WSrep can be used by more heterogeneous user applications which use different rating mechanisms. Now WSrep can not accept negative ratings, because our Cre is a probability, we need to ensure average of user ratings and Cre within an identical data range for computational simplification and data consistency. Taking this into consideration, we need to normalize user ratings into range [0,1] first:

$$Rating = \frac{R}{Max} \tag{1}$$

  Where R denotes the value of rating and Max is the possible maximal value of this rating mechanism. The trustworthy third-party computes the average of ratings in every fixed time window (e.g. e-bay's reputation system [1] where average of ratings can be computed every week, month or six months) and shares them with all customers.
- Cre. The Cre denotes the possibility of service provider advertising QoS information honestly. The computation of Cre is based on the objective feedbacks generated by user agents. After each transaction, the user applications give objective feedbacks on each element of MQ automatically. They compare the real

values of QoS with the values advertised; if they are compliant, the user applications mark the attributes with 1, otherwise, 0. For example, if the user agent finds the real latency is 3s, but the advertised latency is less than 2s, it will rate the latency with 0. This work depends on an assumption:

Assumption 1:
User applications or user agents, in this paper, is trustworthy. That means the objective feedbacks generated by the user applications can not be modified by the customers.

In WSrep, the Cre has two levels:

– *Cre of Attribute (CoA)* expresses the credibility of a single attribute defined in MQ; the CoA is computed in the trustworthy third-party with the objective feedbacks collected from different users.
– *Cre of Web Service (CoWS)* shows the overall credibility of all the attributes defined in MQ, the CoWS is computed at the user local side.

3.2 The Credibility Model

In this subsection, we model the most important parameter in WSrep which is the Cre. Whether the QoS advertisement of a given service is trustworthy is the key influencing factor of services selection. We need a mechanism to objectively predict it, then we can make the rational decisions for customers. Taking this situation into account, an approach based on Bayesian learning theory is needed to model the Cre.

Bayesian learning theory is a statistical theory of making statements about uncertain events. This theory is widely applied in many research fields (i.e. scientific prediction, game-theoretical analyses, decision making and statistics). According to Bayesian learning theory, initially events of interest are assigned a prior belief which reflects existing knowledge about the event and the problem area. Later, as new information (sample) becomes available, the beliefs are updated using the Bayes' rule.

We let a random variable θ denote the CoA. Then we model the CoA using Bayesian learning theory. This work depends on an assumption:

Assumption 2:
For each attribute of MQ, we assume that we have the priors of CoA (results of former evaluations) and represent them as:

$$Priors \equiv \{\theta_1, \theta_2, \theta_3 ......\}$$

We let n denotes the total number of transactions preformed by a given service in a time window and the random variable X denotes the times of compliance of a single attribute among n transactions, which is the sample. So the value of X can be defined as follow:

$$x = \sum_{i=1}^{n} OF_i, i = 1, 2, ..., n \tag{2}$$

Where OF denotes the objective feedbacks generated for this attribute. Obviously, X is binominal distributed ($b(n, \theta)$). According to Bayesian learning theory, a binominal distribution has a Beta-prior ($\pi(\theta) \sim \beta(a,b)$), the resulting posterior is also Beta-distributed ($h(\theta \mid x) \sim \beta(a+x, n+b-x)$). The Beta-distribution is a two parameter distribution whose parameters are denoted by a and b. In order to compute these two parameters, we let $\overline{\theta}$ denotes the average of priors and δ denotes the standard deviation of priors. Otherwise the expectation of $\beta(a,b)$ is $a/(a+b)$ and the standard deviation is $\sqrt{\frac{ab}{(a+b)^2(a+b+1)}}$. So the estimate of a and b are computed according to the formula (3)

$$\begin{cases} \hat{a} = \overline{\theta} \times \left(\frac{(1-\overline{\theta}) \times \overline{\theta}}{\delta^2} - 1\right) \\ \hat{b} = (1-\overline{\theta}) \times \left(\frac{(1-\overline{\theta})\overline{\theta}}{\delta^2} - 1\right) \end{cases} \tag{3}$$

Then we model the CoA as the posterior average estimation of θ (according to Bayesian learning theory, the posterior average estimation of a random variable denotes the most possible average of this variable), which is defined in (4):

$$CoA = \hat{\theta}_E = \frac{a+x}{n+a+b} \tag{4}$$

We assume that a given service has m attributes defined in MQ, the CoWS is defined as weighted average of CoA:

$$CoWS = \sum_{i=1}^{m} \omega_i \times CoA_i \quad , i = 1, 2, \ldots\ldots, m \tag{5}$$

Where ω_i denotes user weight on each attribute of MQ and $\sum_{i=1}^{m} \omega_i = 1$, customers weight each attribute depending on their own need of trust (e.g., a customer may focus on the Cre of latency, so he will weight it more) and the user applications ensure the sum of weighs is 1. In this paper, we do not focus on how the weights are given.

3.3 The Reputation Model

In this subsection, we formalize the parameters introduced above to present the reputation metric. In WSrep, the Reputation of Web Service combines average of user ratings and the CoWS, which is used to measure the level of trust of a service implementation. The WSrep is defined in (6):

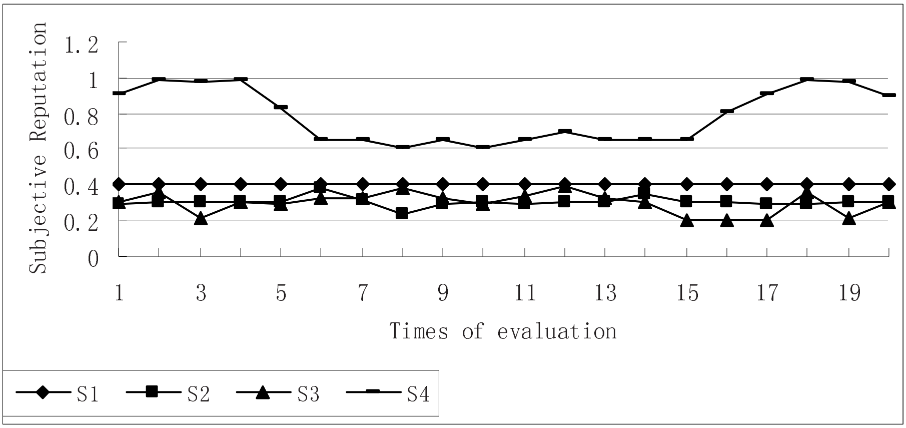
$$WSrep = \alpha \times \frac{\sum_{i=1}^{n} Rating}{n} + \beta \times CoWS \tag{6}$$

Where α and β denote the normalized weight factors for the collective rating and Cre and they need to follow the limitation that the sum of α and β is always 1. The α and β parameters can be used to assign different weights according to different needs of customers. For instance, if a user believes the subjective view of performance history of a service is more reasonable, he can give α a higher value. For a given service, the metric shows that the WSrep focus not only on the subjective view of overall performance of a given service but also the objective view of attributes defined in MQ.

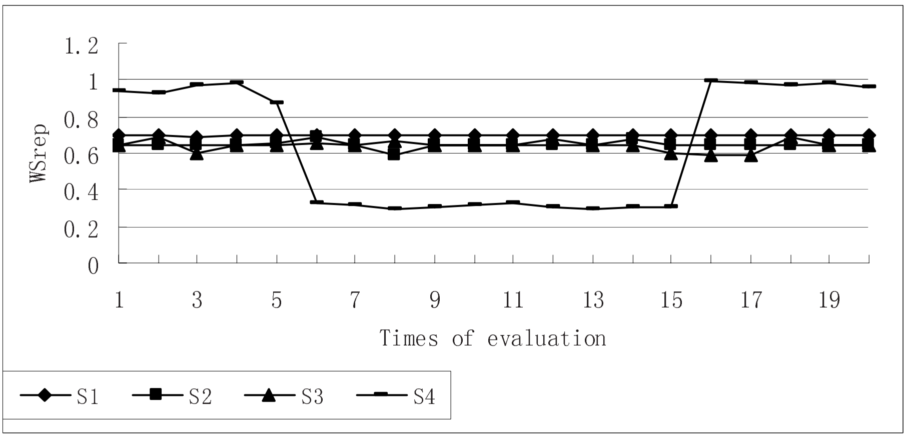
4 Experimentation

We performed two sets of experiments to evaluate the WSrep approach and show its feasibility, effectiveness, and benefits. For comparison, we implemented the WSrep approach and the subjective reputation approach. Further, 4 samples of services will be tested in a services selection scenario.

We divide the services providers into two types, one is honest (always delivering promised QoS), and the other is strategic (colluding with dishonest raters to cheat customers and fool the reputation systems). In our experiments, 4 services are chosen, denoted as {S1, S2, S3, and S4}. The cardinality of MQ of each service is set to be 3 and the user weights on them are generated randomly. For each $q \in MQ$, the priors of Cre are set to be {0.45, 0.50, and 0.55} impartially and α and β are all set to be 0.5. Our experiments also consist of 100 subjective raters (supposed customers and 60% of them are set to be malicious) and corresponding 100 objective raters (supposed software agents). The subjective raters rank services with $i \in \{0,1,2,3,4\}$, and objective raters rank each $q \in MQ$ with 1 or 0. The 4 services we designed have different characteristics, S1 is an excellent service (is honest and offers the best QoS), the honest subjective raters always rank S1 with 4 and the objective feedbacks are always 1; S2 and S3 are two good services (is honest and offers less best QoS). For S2 and S3, each subjective rating given by honest raters is generated randomly from 2 to 4 and the objective feedbacks are always 1; S4 is a strategic service, it colludes with the 60 malicious raters who always rate other 3 services with 0 and rate S4 with 4. When S4 builds it own reputation, each subjective rating given by honest raters is generated randomly from 2 to 4 and the objective feedbacks are always 1; when milking from it, each subjective rating given by honest raters is generated randomly from 0 to 1 the and the objective feedbacks are seldom 1. The subjective reputation and WSrep of the four services are computed 20 times, in each time window, 1000 transactions are performed (each customer does 10 times equally).



a) Values of subjective reputation



b) Values of WSrep

Fig. 1. The benefit of WSrep-based services selection

Fig. 1 shows the values of the reputation of the 4 services using different computational models.

Reputation-based services selection is to choose the honest service which has the highest reputation value depending on once evaluation. For judging the qualities of reputation models used for services selection, we propose a new criterion named Wrong Selection Rate (WSR), which is computed as:

$$WSR = \frac{T_{wrong}}{T_{right} + T_{wrong}} \tag{7}$$

Where T_{wrong} denotes the times of choosing the dishonest services and T_{right} denotes the times of choosing the honest services. The reputation model is better when its WSR is lower.

Fig.1 a) shows the values of subjective reputation of the 4 services. The malicious raters make the S1, S2 and S3 lose their advantages completely (values change from 0.2 to 0.4). This set of experiments also shows S4 milks it reputation from the No.6 time of evaluation to the No. 15 time of evaluation. However, if we use this reputation model to select services, S4 will be chosen at any time (because the reputation values of S4 are higher than other services' at any time). So in our experiments, the WSR of subjective reputation is 100%, which denotes the subjective model of reputation will be disabled under attack of 60% malicious raters.

As expected, Fig.1 b) shows the benefits of WSrep used in the services selection. WSrep makes the reputations of S1, S2, and S3 increase evidently. We admit that S4 will be chosen using WSrep when S4 builds its reputation (because of the malicious rating attack). However, when S4 milks its reputation, the reputations of S1, S2, and S3 exceed S4's. So the WSR of WSrep is 50%. This denotes WSrep is more efficient and robust than subjective reputation model.

5 Related Work

Reputation systems have been studied in several distinct research areas, such as economics, sociology and computer science. In this section, we first review related works in P2P environment, and then review a number of recent works on building reputation systems in web services scope.

Kamvar et al. [13] proposed EigenTrust system for Gnutella like P2P file sharing network. Their work is based on the notion of transitive trust and addressed the collusion problem by assuming some peers can be pretrusted. Their algorithm showed promising results against some threat models. However, the pretrusted peers may not be available in fact and their complex algorithm requires strong coordination of peers. The efficiency of P2P networks will be low if applying such a system. Li Xiong et al. [3] proposed a reputation-based trust supporting framework-PeerTrust which is a combination of five parameters: feedbacks, the number of transactions, the credibility of the feedback sources, transaction context factor and the community context factor. They tried to use the parameter credibility of feedback to find the malicious raters. However, the complex algorithm of credibility made this approach can only be applied in a small community, because if the number of peers is large, differentiating honest raters and dishonest raters is a very heavy work. So we argue that reducing the influence of malicious rating attack is more applicable than discovering each dishonest rater.

The reputation systems researched in web service field are focused on ensuring the providers delivering their promised QoS and helping the customers select the trustworthy services. Most of existing web services reputation models are not strong and flexible enough to reduce threats. Liangzhao Zeng et al. [8] modeled the reputation of web service as average of user ratings. E. Michael Maximilien et al.[6] introduce a conceptual model of web service reputation, they denote within a specific domain the reputation of the service depends on the subjective view of the users of the service on the various attributes. They also present a UML static model for the components that make up the reputation of a service. The main shortcoming of these reputation model is that the rating-based or subjective view-based reputation is not sophisticated enough to deal with malicious rating attacks so that these systems are easy to be disabled. Interestingly, Sravanthi Kalepu et al. [10] observed the importance of the objective view of services' performance history. They modeled reputation of web service as f (User Ranking, Compliance, and Verity), where compliance refers to the service provider's ability to meet the service level of each QoS parameter laid out in the SLA without incurring penalties and verity (the objective factor the authors advocated) is a mathematical variance represents the compliance levels. However, the computation of verity is based on the subjective user rankings, so it is not "real" objective. If malicious raters exist, the verity will not make any sense too. Further, above reputation models all express the global view of trustworthiness of a given service, it can not satisfy the customers who have special trust needs when they face the services selection problem.

Our work differs from them in a number of ways. First, we emphasize the objective factor is as important as the subjective factor in web services' reputation systems. Then we introduce Cre to reduce the negative influence of the malicious rating attack. The Cre is computed based on objective feedbacks generated automatically by customers' agents, so this value is trustworthy. Second, we integrate the average of user ratings and Cre into the reputation model. A given customer can customize the values of reputation by assigning weights. The personal reputation is more valuable than the general reputation in services selection processes. Third, we run a series simulation-based experiment to show the effectiveness and benefits of our approach.

6 Conclusion

Web services environments offer both opportunities and threats. Building a flexible and robust reputation system is the most efficient way to minimize threats. In this paper, we have described WSrep-a novel reputation model used for selecting the most trustworthy web service. The WSrep model is made up of subjective user ratings and credibility of QoS advertisements which is an objective factor to predict whether the providers delivering their promised QoS honestly. For modeling the credibility, we first proposed a customer measurable QoS model used by software agents to validate the compliance of the actual QoS values and their advertisements automatically and generate the objective feedbacks. Then the credibility is modeled based on the Bayesian learning theory. Finally, we reported initial simulation-based experiments, demonstrating the effectiveness and benefits of our approach.

References

1. ebay, http://www.ebay.com, 2005.
2. bindingpoint, http://www.bindingpoint.com/, 2005.
3. Li Xiong, and Ling Liu. PeerTrust: Supporting Reputation-Based Trust for Peer-to-Peer Electronic Communities. IEEE Transaction on Knowledge and Data Engineering. Vol.16, No.7, JULY 2004.
4. Shuping Ran. A Model for Web Services Discovery With QoS. ACM SIGecom Exchanges. Vol.4, No.1, 2003.

5. E. Michael Maximilien, and Munindar P. Singh. Reputation and Endorsement for Web Services. ACM SIGecom Exchanges. Vol.3, No.1, Dec 2001.
6. E. Michael Maximilien, and Munindar P. Singh. Conceptual Model of Web Service Reputation. ACM SIGMOD Record. Vol. 31, No. 4, Dec 2002.
7. E. Michael Maximilien, and Munindar P. Singh. Toward Autonomic Web Services Trust and Selection. In *ICSOC'04,* November 15–19, 2004.
8. Liangzhao Zeng, Boualem Benatallah, Marlon Dumas, Jayant Kalagnanam, and Quan Z. Sheng. Quality Driven Web Services Composition. In *Proceedings of international World Wide Web conference* 2003.
9. Dimitris Gouscos, Manolis Kalikakis, and Panagiotis Georgiadis. An Approach to Modeling Web Service QoS and Provision Price. In *Proceedings of the Fourth International Conference on Web Information Systems Engineering Workshops* (WISEW'03).
10. 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).
11. Daniel A. Menasce. QoS Issues in Web Services. IEEE INTERNET COMPUTING, 2002.
12. QoS for Web Services: Requirements and Possible Approaches. http://www.w3c.or.kr/kr-office/TR/2003/ws-qos/.
13. S.Kamvar, M.Scholsser, and H.Garcia-Molina, The EigenTrust Algorithm for Reputation Management in P2P Metworks. In *Proceedings of international World Wide Web conference* 2003.