Trust Assessment for Web Services under Uncertainty

Zaki Malik[1] and Brahim Medjahed[2]

[1] Department of Computer Science, Wayne State University, MI. 48202
zaki@wayne.edu
[2] Department of Computer Science, University of Michigan-Dearborn, MI. 48120
brahim@umd.umich.edu

Abstract. We introduce a model for assessing the trust of providers in a service-oriented environment. Our model is cooperative in nature, such that Web services share their experiences of the service providers with their peers through ratings. The different ratings are aggregated using the "statistical cloud model" defined for uncertain situations. The model can uniformly describe the concepts of randomness, fuzziness, and their relationship in quantitative terms. By incorporating the credibility values of service raters in the model, we can assess a service provider's trust. Experiment results show that our proposed model performs in a fairly accurate manner.

1 Introduction

With the introduction of *Web services*, applications can now be automatically invoked by other Web clients. A Web service is a self-describing software application that can be advertised, located, and used across the Web using a set of standards (such as WSDL, UDDI, and SOAP) [29]. Businesses are increasingly using Web services to automate interactions both with their customers (B2C) and amongst each other (B2B). It is expected that future Web enterprises would exhibit a loose coupling of smaller applications offered by autonomous providers [26][29]. A primary goal of the Web services technology is therefore enabling the use of Web services as independent components in Web enterprises, that are automatically (i.e., without human intervention) formed as a result of consumer demand and which may dissolve post demand-completion [26].

Automatic Web services interactions entail that Web services have to determine to which extent they may *trust* other services to provide the required functionality, before they interact with them [20]. By definition, Web services are autonomous (i.e., provided by independent service providers), highly volatile (i.e., low reliability), and *a priori* unknown (i.e., new or no prior history) [29]. As a plethora of Web services are expected to compete in offering similar functionalities, a key requirement is then to provide mechanisms for the quality access and retrieval of services [25] [29]. Web services may make promises about the

P.P. Maglio et al. (Eds.): ICSOC 2010, LNCS 6470, pp. 471–485, 2010.

provided service and its associated quality but may fail partially or fully to deliver on these promises bringing down the quality of the whole enterprise. Thus, the challenge lies in providing a framework for enabling the selection and composition of Web services based on trust parameters. The rationale behind the need for trust is the necessity to interact with unknown entities that have varied quality delivery levels [2]. There is a growing consensus that the Web service 'revolution' would not eventuate until trust related issues are resolved [4].

Trust has been defined as "an assured reliance on the character, ability, or strength of someone or something." Establishing trust is therefore a precondition for any transaction [2][22]. In a service-oriented environment, trust correlates to the ability of a service to perform the required functionality in an acceptable manner. The inherent open and large-scale nature of Web services means that traditional security approaches as confidentiality, authentication, authorization, etc. are insufficient for completely instilling trust. For instance, a provider's authentication or authorization credentials cannot guarantee that it will exercise these privileges in an expected manner [19]. When interacting with unknown providers, service consumers are thus usually interested in gaging provider reliability in delivering the required functionality (on top of traditional security mechanisms). Research results show that such a trust assessment process is facilitated by incorporating the "wisdom of crowds" through reputation ratings and recommendations [8] [20]. For example, several studies attribute eBay's commercial success to its reputation mechanism, known as eBay's Feedback Forum which has been effective in deterring dishonest behavior, and stimulating eBay's growth [30] [10]. Similar studies have investigated and generally confirmed that reputation systems benefit both sellers and buyers in e-auctions[15]. Reputation is defined as the confidence in the ability of a specific provider to fulfill a certain task [20]. It is a subjective assessment of a characteristic or an attribute ascribed to one entity by another based on observations or past experiences. Normally experiences from more than one source are assimilated to derive the reputation. This increases the subjectivity of trust and creates uncertainty.

In recent years, theoretical and experimental research has explored the subjective nature of trust. These works are primarily rooted in probability theory, evidence/belief models, or fuzzy logic. Probability based models usually do not consider the element of fuzziness in building trust [3] [34]. Since the reasoning is done in a purely statistical manner, they tend over-formalize trust's subjectiveness. For example, Bayesian systems take binary ratings as input and assess trust through updating of the beta probability density function [38] [33]. This process is fairly complex to comprehend and implement, and loses the component of fuzziness inherent in trust assessment. Models based on evidence and belief theory exhibit similar characteristics with added complexity [13] [34]. On the other hand, fuzzy logic based systems use precise set memberships for defining fuzziness of subjective trust. However, these solutions fail to consider the randomness and uncertainty of membership in those fuzzy sets [9] [27]. We propose a solution that incorporates uncertainty and fuzziness of trust to provide a more unified and holistic assessment. Our model employs the statistical cloud

model which defines a way for modeling the transition between a linguistic term of a qualitative concept and its quantitative representation under uncertain and fuzzy conditions.

The paper is organized as follows. In Section 2, we provide an overview of a statistical model for predicting values in uncertain situations. In Section 3, we extend this model to evaluate trust of service providers. Section 4 provides experiment results, and verifies the applicability of our proposed model. Section 5 provides a brief overview of some related work, while Section 6 concludes the paper.

2 Statistical Cloud Model

The basis of the statistical cloud model (or simply, the cloud model) is that fuzziness and randomness are complementary and essentially inseparable concepts when considered in linguistic terms. It states that the concept of fuzzy membership functions is not sufficient for representing the uncertainty and imprecision in real world settings, and probability theory needs to be incorporated to overcome this inadequacy. In essence, a cloud model can uniformly describe the concepts of randomness, fuzziness, and their relationship in quantitative terms. Experiment results have shown that the cloud model exhibits higher levels of simplicity and robustness in comparison with traditional fuzzy logic and probability based methods [17] [18]. In the following, we provide a brief overview of the cloud model.

Let U be the quantitative universe of discourse, and C denote a qualitative concept associated with U. If $x \in U$ is a random realization of C, and $\mu(x) \in [0, 1]$ is a random variable with stable tendency denoting the degree of certainty for x belonging to C, that is:

$$\mu: U[0, 1] \qquad \forall x \in U \qquad x \rightarrow \mu(x)$$

The distribution of x in U is called the cloud (denoted $C(X)$) and each x is called a cloud drop. Note that in probabilistic terms, $x \in U$ is not a simple random number but it has a certainty degree, which itself is also random and not a fixed number. The cloud is composed of a number of drops, which are not necessarily ordered. The underlying character of the qualitative concept is expressed through all cloud drops. Hence the overall feature of the concept is more precisely represented by a large number of drops. The certainty degree of each cloud drop defines the extent to which the drop can represent the concept accurately. Formally, a cloud's quantitative representation is defined over a set of N ordered pairs (x_i, y_i), where x_i is a cloud drop, and y_i is its certainty degree, with $1 \leq i \leq N$.

A one-dimension normal cloud model's qualitative representation can be represented by a triple of quantitative characteristics: Expected value (Ex), Entropy (En) and Hyper-Entropy (He). Ex is the expectation of the cloud drops' distribution, i.e., it corresponds to the center of gravity of the cloud (containing elements fully compatible with the qualitative concept). En represents the

uncertainty measurement of a qualitative concept. It is determined by both the randomness and fuzziness of the concept. En indicates how many elements could be accepted to the qualitative linguistic concept. He is a measure of the dispersion on the cloud drops. It can also be considered as En's uncertainty. Vector $\boldsymbol{v} = (Ex, En, He)$ is called the eigenvector of a cloud [17].

The transformation of a qualitative concept expressed by Ex, En, and He to a quantitative representation expressed by the set of numerical cloud drops is performed by the forward cloud generator [18]. Given these three digital characteristics (Ex, En, He), and the number of cloud drops to be generated (N), the forward cloud generator can create these N cloud drops in the data space with a certainty degree for each drop that each drop can represent the qualitative concept. The procedure is:

1. Generate a normally distributed random number F with mean En and standard deviation He.
2. Generate a normally distributed random number x with mean Ex and standard deviation F.
3. Calculate $y = e^{-\frac{(x-Ex)^2}{2(F)^2}}$.
4. (x, y) represents a cloud drop in the universe of discourse.
5. Repeat Steps 1-4 until N cloud drops are generated.

Figure 1(a) shows the graph of a one-dimensional cloud whose digital characteristics are (0.7, 0.1, 0.01). A similar cloud with same Ex and En, but a different He (0.7, 0.1, 0.5) is shown in Figure 1(b). As defined in the above algorithm, the quantitative value of cloud drops is determined by the standard normal form distribution function. Hence, the certainty degree function adopts a bell-shaped curve. This is similar to the one adopted in fuzzy set theory. As mentioned earlier, the normal cloud model is therefore an inclusive model based on probability theory and fuzzy set theory, and is able to depict randomness in the former and fuzziness in the latter.

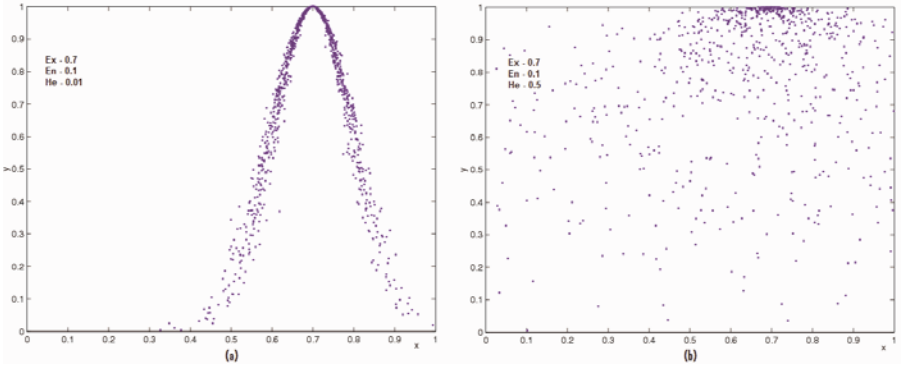

Fig. 1. Normal Cloud with Same Ex and En, but Different He Values

3 Statistical Cloud-Based Trust Model

We propose a trust model that is distributed in nature. In contrast to third-party-based traditional approaches for trust management, no single entity is responsible for collecting, updating, and disseminating ratings provided by different consumers. Each service consumer records its own perceptions of the reputation of only the services it actually invokes. This perception is called personal evaluation. For each service s_j that it has invoked, a service consumer t_x maintains a p-element vector $PerEval_j^x$ representing t_x's perception of s_j's behavior. Different strategies may be adopted in updating $PerEval_j^x$. A simple one may be a *per-invocation* update. Upon an invocation of service s_j, the delivered quality $QRef_d$ is compared to service s_j's promised quality $QRef_p$ and, if necessary, a trust updating algorithm is run to compute the new personal evaluation of service s_j. In essence, personal evaluation reflects the *Quality* performance of the provider in consumer's views. The personal evaluation $PerEval_j^x$, represents only consumer t_x's perception of the provider s_j's reputation. Other service consumers may differ or concur with t_x's observation of s_j. A service consumer that inquires about the reputation of a given service provider from its peers may get various differing personal evaluation "feedbacks." To get a correct assessment of the service provider's behavior, all the personal evaluations for s_j need to be aggregated. Assume L denotes the set of service consumers which have interacted with s_j in the past and are willing to share their personal evaluations of s_j. We assume that L is not empty, i.e., some service willing to share information can be found. Thus, $L \subseteq T$ with $L \neq \emptyset$ and each service x in L has $PerEval_j^x$ values for s_j. Then, consumer x's trust over s_j's ability to deliver is defined as:

$$Trust(s_j) = \bigwedge_{x \in L} (PerEval_j^x) \tag{1}$$

where $\bigwedge$ represents the aggregation function. Equation 1 provides a first approximation of how the trust may be assessed. However, it involves various factors that need to be precisely defined and measured.

The foremost drawback of feedback-only based systems is that all ratings are assumed to be honest and unbiased. However, in the real world we clearly distinguish between the testimonies of our sources and weigh the "trusted" ones more than others [36]. A Web service that provides satisfactory service (in accordance with its promised quality ($QRef_p$)), may get incorrect or false ratings from different evaluators due to several malicious motives. In order to cater for such "bad-mouthing" or collusion possibilities, a trust framework should weigh the ratings of highly credible raters more than consumers with low credibilities [7] [39] [20]. In our model, the final trust value is calculated according to the credibility scores of the raters (used as the weight).

After each interaction with the provider, apart from rating the provider s_j, the service consumer also updates the credibility of the raters that provided a rating for s_j. The service consumer computes the Euclidean distance (d) between the consumer's own experience (OE) and the provided rating (V_i). If d is less than a pre-defined threshold (δ), the credibility is increased in a linear

manner. Otherwise, the rater's credibility is decreased exponentially by a factor of d, i.e., greater the d, more the rater credibility will decrease. This is in accordance with the sociological trust building process where it is difficult to gain inter-personal trust, but easy to lose it [6]. Since all transactions may not be equally weighed in terms of their importance, a service consumer may decide to decrease a dishonest rater's credibility according to the transaction's "impact". The transaction impact factor (τ) lies in the range [0, 1] and is assigned a low value for high impact transactions, and vice versa by the service consumer. The general formula for rater credibility is thus:

$$Cr^t = \begin{cases} Cr^{t-1} + c(\delta - d) & \text{if } d \leq \delta; \\ Cr^{t-1} \times \mathrm{e}^{-(d+\tau)} & \text{otherwise.} \end{cases}$$

where c is the linear increment factor, weighted by the difference between δ and d. This implies that a lower value of δ will cause a lower increment in the value of c and hence the rater's credibility. Cr^t is the new credibility, and Cr^{t-1} is the rater's previous credibility value. The credibility of a service rater lies in the interval [0, 1] with 0 identifying a completely dishonest rater and 1 an honest one. In cases where no Cr^{t-1} exists, i.e., the rater and consumer have not previously interacted, the rater's initial credibility is set at the middle (0.5) to indicate impartiality. However, previous research has shown that assigning predefined high or average values may encourage "reputation white-washing" [21]. Therefore, we dampen the bootstrap value by weighing in the consumer's pessimistic/optimistic preferences towards services interactions, i.e.,

$$Cr_{bootstrap} = 0.5 \times \lambda$$

where λ denotes the consumer's pessimistic/optimistic preference in the range [0, 1]. A high λ value indicates an optimistic consumer, one that is willing to trust the testimony of a new rater. Alternatively, $\lambda \leq 0.5$ indicates a pessimistic consumer. The choice of λ is at the discretion of the service consumer. However, to provide a better estimate of the consumer's propensity to accept, we set λ as the *ratio* of the total number of times the ratings submissions (by all raters) are deemed useful (k) by the service consumer, over the total number of rating submissions received by the service consumer (n). This is similar to the manner in which peer recommendations are evaluated for usefulness in "recommender systems" [16][35]. The λ factor is:

$$\lambda = \frac{\sum_{i=1}^{k} U_i}{\sum_{x=1}^{n} V_x} \tag{2}$$

where U_i is the submission where the rater was termed honest (i.e., $d \leq \delta$) and V_x denotes the total number of rating submissions.

Reputation information of a service provider decays with time [20], [23]. Hence all the past reputation data may be of little or no importance. For instance, a Web service performing inconsistently in the past may ameliorate its behavior. Alternatively, a service's performance may degrade over time. It may be the

case that considering all historical data may provide incorrect reputation scores. In order to counter such discrepancies, we incorporate temporal sensitivity in our proposed model. The rating submissions are time-stamped to assign more weight to recent observations and less to older ones. This is termed as "reputation fading" where older perceptions gradually *fade* and fresh ones take their place. We adjust the value of the ratings as:

$$PerEval_j^{x:t} = PerEval_j^{x:t-1} * f_d \tag{3}$$

where $PerEvall_j^x$ is as defined above and f_d is the reputation fader. In our model, the recent most rating has the fader value 1 while older observations are decremented for each time interval passed. When $f_d = 0$, the consumer's rating is not considered as it is outdated. The "time interval" is an assigned factor, which could be anywhere from a single reputation inquiry, ten inquiries or even more than that. All inquiries that are grouped in one time interval are assigned the same fader value. In this way, the service consumer can define its own temporal sensitivity degree. For example, a service can omit the fader value's effect altogether by assigning it a null value. We propose to use a fader value that can then be calculated as: $f_d = \frac{1}{\sqrt{P_u}}$, where P_u is the time interval difference between the present time and the time in which the rating was collected from the rater. This allows the convergence of reputation to a very small value as time passes. Note that the consumer can assign a group of ratings collected at different times to have the same time-stamp, and hence lie in the same time interval. As mentioned earlier, other calculated values for the fader are also acceptable.

Characteristics Extraction

The backward cloud generator allows transformation of the cloud model from its quantitative representation to a qualitative one. We incorporate rater credibility values and majority rating to produce the three digital characteristics of the cloud (Ex, En, He). Given a set of N ratings $PerEval_j^x (x = 1, 2, ..., N)$, we can extract the three characteristics as:

1. Update $PerEval_j^x$ values using f_d, for all ratings (including previous time instances).
2. For each rater x, update Cr_x (using equations defined previously).
3. Calculate
$$Ex = \frac{\sum_{x=1}^{N}(Cr_x PerEval_j^x)}{\sum_{x=1}^{N} Cr_x}$$
4. Calculate
$$En = \sqrt{\frac{\pi}{2}} \times \frac{\sum_{x=1}^{N} Cr_x |PerEval_j^x - Ex|}{\sum_{x=1}^{N} Cr_x}$$
5. Calculate
$$He = \sqrt{\frac{\sum_{x=1}^{N} Cr_x (PerEval_j^x - Ex)^2}{\frac{(N'-1)\sum_{x=1}^{N} Cr_x}{N'}} - (En)^2}$$
where N' is the number of non-zero credibilities.

Trust Decision

The next step is using the three discovered characteristics to make a subjective assessment of the provider's trust. Since He is a measure of En's uncertainty, we only use Ex and He to quantify the provider's trust and the associated uncertainty. This allows us to consider the latest majority view of the provider's reputation and the decentralization of ratings from it. A higher value of Ex therefore indicates high reputation, while a small He indicates the stability of the ratings around this decision. Intuitively this makes sense, but for a large N, making these comparisons is non-trivial. For instance, Ex and He can occur together in one of four forms: one is high/low the other is low/high, both are high, or both are low. Therefore, to quantify the relationship between the two characteristics, i.e., the provider (s_j)'s trust assessment, we use:

$$Trust(s_j) = \begin{cases} 1 - \frac{He}{Ex+He} & \text{if } Ex \neq 0 \ \& \ He \neq 0; \\ Ex & \text{if He} = 0; \\ 0 & \text{if Ex} = 0; \end{cases}$$

where both $Ex \neq 0$ and $He \neq 0$.

4 Experiments

We have performed a number of experiments to show the applicability of the proposed statistical cloud-based trust model. We used Matlab for simulating the services interactions and ratings. The environment consists of five service providers and twenty service consumers (who act as raters) that interact over a period of twenty iterations. At each iteration, the raters report their past experiences (from the previous iteration) of the service providers. For simplicity, each rater interacts with all the service providers in each time instance. Therefore, the fader value (f_d) is set to 1.

The five service providers exhibit different dynamic behaviors. The first provider behaves consistently with high trust values, i.e., it behaves rationally and does not engage in any malicious activity. The next provider performs with consistently low trust values. It represents providers that are always looking to take advantage of the consumer. The third provider performs with high values for the first 10 iterations but then suffers a performance degradation. This strategic provider aims to build a reputation by initially performing honestly, and then starts "milking" [39] the attained reputation. The fourth provider acts in an opposite manner to the third provider where it performs with low values in the beginning. After the 10th iteration, it ameliorates its behavior and starts performing with high trust values. This provider represents the class of providers that learn from their mistakes. The provider performs in a random manner, oscillating between high (performing as promised) and low trust values (acting maliciously).

The service raters are distinguished into two classes: honest and dishonest raters. An honest rater provides the trust value it experiences, but a dishonest

rater generates a rating that differs at least by 0.3 points from the actual rating. Say the provider's trust value was 0.9, then a dishonest rater would generate a value between [0.1 and 0.59]. These two classes of raters can be related to each other in one of three ways in the environment: the number of honest raters can exceed those of dishonest raters, honest and dishonest raters can be equal in number, or dishonest raters can out-number honest raters. We set the inequalities in rater behaviors (first and third scenario) to be significant (an 80-20 ratio imbalance is used). The different classes of raters and five provider behaviors mentioned above (and any combination thereof) cover any behavior that a service may exhibit. This ensures that the experiment samples are representative of the real world environment which contains a variety of provider and rater behaviors.

In the first experiment, honest raters (ones with high credibility values) out-number dishonest raters, i.e., 80% of the raters are honest. Figure 2(a) shows the effect of this inequality in calculating the trust value for a provider that exhibits low values in a consistent manner. The dashed-line represents the trust value of the provider estimated using our proposed model, whereas the straight line represents the actual behavior experienced by the trust evaluator. It can be seen that due to the high number of honest ratings, the estimated trust value is almost equal to the actual provider behavior. The small variation in estimated and actual trust is due to the inconsistency brought in by the differences in opinions of credible raters and malicious attempts of non-credible raters. Figure 2(b) shows the case for the same provider (exhibiting same actual behavior) for the case where a large majority of raters are dishonest. In this case, the system "catches up" with dishonest testimonies after a few initial iterations. However, once this learning process is complete, the presence of dishonest raters is minimized and actual vs. estimated trust values become very close. Other classes of providers exhibit similar results, thus we omit those graphs (and associated discussions here).

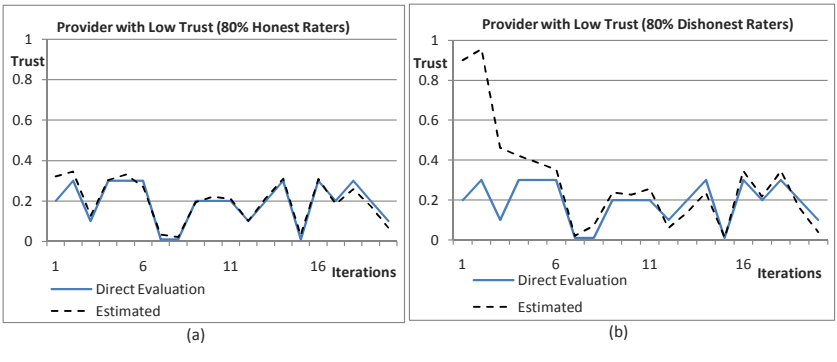

Fig. 2. Service Provider with Consistently Low Trust Values

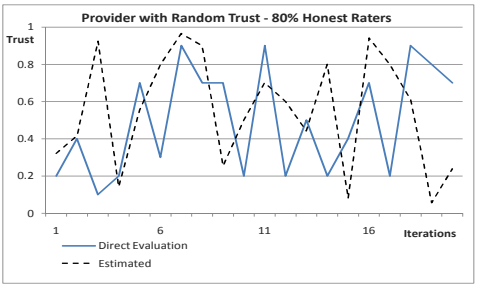

Fig. 3. Service Provider with Random Trust Values: 80% Raters are Honest

Figure 3 shows the trust evaluation process for a provider that exhibits random behavior in terms of trust values. Although the number of honest raters exceeds that of dishonest ones, the system still exhibits inconsistent results. This is due to the manner in which rater credibilities are evaluated and weighed. An honest rater that reports its experienced trust may have its credibility reduced in the next iteration due to the inconsistent behavior of the provider. Consequently, widening the gap between estimated and experienced trust.

We evaluate the error percentage of the proposed model by comparing the output (service provider chosen) against the "actual best service provider alternative" available. Note that since these are controlled experiments, we can identify the best providers for each iteration. A graph depicting these behaviors for all service interactions becomes convoluted. Thus, for brevity Figure 4(b) only shows a snapshot of actual service provider behaviors for interaction numbers 9 through 13. We can see that Service Provider 1 shows consistently high trust values, while Provider 3's performance (and hence trust) drops at 11th. iteration onwards. In contrast, Provider 4's trust values shift to higher values at the same point. Similarly, Provider 2 performs with consistently low values, while Provider 5 exhibits random behavior switching between high and low trust values. Figure 4(a) shows the trust values estimated using the proposed model for the same time period. The model chooses the service provider with the highest trust value at each iteration for interaction. For instance, at iteration 10, Service Provider 3 is estimated to be the "best" service available, while at iteration number 11 Provider 1 is chosen and Provider 5 is chosen at iteration 12 (and so on).

Table 1 shows all iterations, best provider alternatives for each iteration, and the service provider chosen by our model. The final column labeled "Error" places an `X` if the chosen service is incorrect. There are six instances in which our model chose an incorrect service as per the given data. Apart from iterations 3 and 4, the other four errors were reported because the model chose Provider 5. Note that this class of providers behaves randomly at each iteration. Thus, it is difficult for the system to predict or report correct trust values. However, such provider behavior is expected to be rare [20]. The errors of iterations 3 and 4 can be termed as "minor" since both Provider 1 and 3 have high trust values

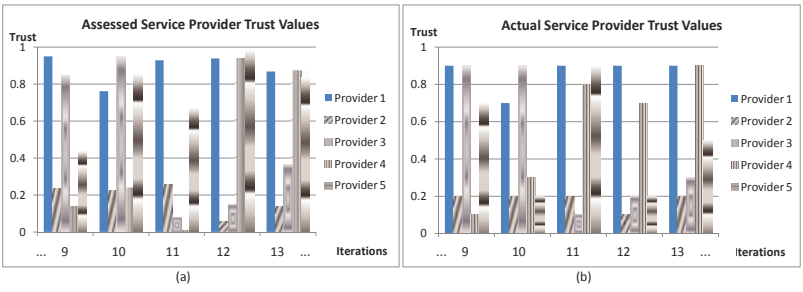

Fig. 4. Actual Service Provider Trust Values Compared for Iterations 9 through 13

Table 1. Error Evaluation of the Proposed Model

Iteration	Best Service Provider Available	Chosen Service Provider	Error
1	P1 or P3	P5	X
2	P1 or P3	P1	-
3	P3	P1	X
4	P1	P3	X
5	P1, P3 or P5	P3	-
6	P1 or P3	P3	-
7	P1, P3 or P5	P1	-
8	P1 or P3	P5	X
9	P1 or P3	P1	-
10	P3	P3	-
11	P1 or P5	P1	-
12	P1	P5	X
13	P1 or P4	P1	-
14	P4	P4	-
15	P4	P4	-
16	P1	P1	-
17	P4	P5	X
18	P1	P1	-
19	P4	P4	-
20	P1	P1	-

(note that Provider 3 switches behavior at the 11th. iteration). If we consider *all* reported errors, we can say the proposed model has 70% accuracy, while if we consider P1 and P3 comparable (which they are), then system accuracy jumps to 80%. Note that, these error evaluation results are for the case where 80% of the raters are dishonest. When 80% of the raters are honest, error percentage is reduced to less than 5%. In light of these results, we can conclude that our proposed model estimates a service provider's trust in a fairly accurate manner even under uncertainty.

5 Related Work

Trust assessment involves several components, including modeling, data collection, data storage, communication, assessment, and safeguards. Over the years, several research initiatives have worked on most of these problems. Similar to our model, most initiatives equate trust with reputation, i.e., the higher the reputation of a provider, the more trustworthy it is, and vice versa. Varied disciplines including economics, computer science, marketing, politics, sociology, and psychology have studied reputation-based trust in several contexts [8]. In the recent past, these research activities have gained momentum. In computer science, reputation has been studied both in theoretical areas and practical applications. Theoretical areas where reputation has been studied include game theory [12], Bayesian networks [37], overlay networks,[32] and social networks [6] to name a few. Theoretical literature that addressed reputation focused on proving properties of systems based on reputation. For example, results from game theory demonstrate that there are inherent limitations to the effectiveness of reputation systems when participants are allowed to start over with new names [31]. In [11], the authors study the dynamics of reputation, i.e., growth, decay, oscillation, and equilibria. Practical literature on reputation is mainly concerned with the applications of reputations. Major applications where reputation has been effectively used include e-business, peer-to-peer (P2P) networks, grid computing systems [1], multi-agent systems [33], Web search engines, and ad-hoc network routing [5]. In the following, we give a brief overview of a few reputation management frameworks for P2P systems and Web services since these are closely related to our research.

PeerTrust [39] is a P2P reputation management framework used to quantify and compare the trustworthiness of peers. In PeerTrust, the authors have proposed to decouple feedback trust from service trust, which is similar to the approach undertaken in this paper. Similarly, it is argued that peers use a similarity measure to weigh opinions of those peers highly who have provided similar ratings for a common set of past partners. However, this may not be feasible for large P2P systems, where finding a statistically significant set of such past partners is likely to be difficult. Consequently, peers will often have to make selection choices for peers which have no common information in the system.

In [14], the *EigenTrust* system is presented, which computes and publishes a global reputation rating for each node in a network using an algorithm similar to Google's *PageRank* [28]. Each peer is associated with a global trust value that reflects the experiences of all the peers in the network with that peer. EigenTrust centers around the notion of transitive trust, where feedback trust and service trust are coupled together. Peers that are deemed honest in resource sharing are also considered credible sources of ratings information. This is in contrast with our approach and we feel this approach may not be accurate. Moreover, the proposed algorithm is complex and requires strong coordination between the peers. A major limitation of EigenTrust is that it assumes existence of pretrusted peers in the network.

PowerTrust [40] is a "distributed version" of EigenTrust. It states that the relationship between users and feedbacks on eBay follow a Power-law distribution.

It exploits the observation that most feedback comes from few "power" nodes to construct a robust and scalable trust modeling scheme. In PowerTrust, nodes rate each interaction and compute local trust values. These values are then aggregated to evaluate global trust through random walks in the system. Once power nodes are identified, these are used in a subsequent look-ahead random walk that is based on Markov chain to update the global trust values. Power nodes are used to assess the reputation of providers in a "system-wide absolute" manner. This is in contrast with our approach where each consumer maintains control over the aggregation of ratings to define a provider's reputation. Moreover, PowerTrust requires a structured overlay (for DHT), and the algorithms are dependent on this architecture. In contrast, service-oriented environments or the Web in general do not exhibit such structure.

Despite the abundance in reputation-related literature, little research has focused on the reputation of Web services. In [24], a distributed model for Web service reputation is presented. The model enables a service's clients to use their past interactions with that service to improve future decisions. It also enables services' clients to share their experience from past interactions with Web services. Agents are associated with each Web service, that act as proxies to collect information on and build a reputation of a Web service. The authors present an approach that provides a conceptual model for reputation that captures the semantics of attributes. The semantics includes characteristics, which describe how a given attribute contributes to the overall rating of a service provider and how its contribution decays over time. A similar reputation-based model using a node's first hand interaction experience is presented in [32]. The goal of the model is to increase/maintain QoS values in *selfish* overlay networks. The authors show that in presence of a reputation management system, an overlay network discourages *selfish* nodes. This increases the QoS guarantees in the network. The proposed model considers a node's first hand interaction experience and peer testimonials for deriving node reputations. In this regard, the reputation building process in [32] is similar to our approach. However, the proposed reputation model may not be completely robust and may not provide accurate results. First, the individual experience takes time to evolve over repeated interactions. Second, no distinction is made between the node's service credibility in satisfying consumer requests and its rating credibility. It may be the case that a node performs satisfactorily but does not provide authentic testimonials. We provide an extensive mechanism to overcome these and similar inadequacies.

6 Conclusion

We have presented a trust assessment model for Web services. We focused on an environment where Web services can act as both consumers (i.e., requesters) and providers of services, without the need of a *trusted third party*. This similarity with P2P systems, wireless networks, etc. means that the model is extensible and can be deployed in other contexts. We have also conducted extensive simulations to verify the proposed model. Results exhibit strong evidence that our approach

provides a fairly accurate assessment of provider trust. In the future, we intend to implement the model in a real Web services environment. We also aim to extend the model to include service *compositions*.

References

1. Azzedin, F., Maheswaran, M.: Evolving and Managing Trust in Grid Cmputing Systems. In: Proc. of the IEEE Canadian Conference on Electrical and Computer Engineering, pp. 1424–1429 (May 2002)
2. Bertino, E., Ferrari, E., Squicciarini, A.C.: Trust-X: A Peer-to-Peer Framework for Trust Establishment. IEEE TKDE 16(7), 827–842 (2004)
3. Bharadwaj, K., Al-Shamri, M.: Fuzzy computational models for trust and reputation systems. Electron. Commer. Rec. Appl. 8(1), 37–47 (2009)
4. Birman, K.: The untrustworthy web services revolution. IEEE Computer 39(2), 113–115 (2006)
5. Buchegger, S., Le Boudec, J.-Y.: Performance Analysis of the CONFIDANT Protocol. In: Proc. of the 3rd ACM Intl. Symposium on Mobile Ad Hoc Networking and Computing, June 9-11, pp. 226–236 (2002)
6. Buskens, V.: Social Networks and the Effect of Reputation on Cooperation. In: Proc. of the 6th Intl. Conf. on Social Dilemmas (1998)
7. Delgado, J., Ishii, N.: Memory-Based Weighted-Majority Prediction for Recommender Systems. In: ACM Workshop on Recommender Systems (1999)
8. Dellarocas, C.: The Digitalization of Word-of-Mouth: Promise and Challeges of Online Feedback Mechanisms. In: Management Science (October 2003)
9. He, R., Niu, J., Yuan, M., Hu, J.: A novel cloud-based trust model for pervasive computing. In: International Conference on Computer and Information Technology, pp. 693–700 (2004)
10. Houser, D., Wooders, J.: Reputation in Auctions: Theory, and Evidence from eBay. Journal of Economics and Management Strategy (2005)
11. Huberman, B.A., Wu, F.: The Dynamics of Reputations. TR, Hewlett-Packard Laboratories and Stanford University (January 2003)
12. IBM. Aglet software development kit (2000), `http://www.trl.ibm.com/aglets`
13. Jøsang, A.: A logic for uncertain probabilities. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 9(3), 279–311 (2001)
14. Kamvar, S.D., Schlosser, M.T., Garcia-Molina, H.: The eigentrust algorithm for reputation management in p2p networks. In: Proceedings of the Twelfth International World Wide Web Conference (WWW) (2003)
15. Kesler, C.: Experimental Games for the Design of Reputation Management Systems. IBM Systems Journal 42(3) (2003)
16. Lam, S., Riedl, J.: Shilling Recommender Systems for Fun and Profit. In: Proc. of the 13th International World Wide Web Conference (WWW), New York, NY, USA, pp. 393–402 (2004)
17. Li, D., Han, J., Shi, X., Chan, M.: Knowledge representation and discovery based on linguistic atoms. Knowledge-Based Systems 10(7), 431–440 (1998), KDD: Techniques and Applications
18. Li, D., Liu, C., Gan, W.: A new cognitive model: Cloud model. Int. J. Intell. Syst. 24(3), 357–375 (2009)
19. Malik, Z., Bouguettaya, A.: Rater Credibility Assessment in Web Services Interactions. World Wide Web Journal 12(1), 3–25 (2009)
20. Malik, Z., Bouguettaya, A.: Reputation-based Trust Management for Service-Oriented Environments. VLDB Journal 18(4), 885–911 (2009)

21. Malik, Z., Bouguettaya, A.: Reputation Bootstrapping for Trust Establishment among Web Services. IEEE Internet Computing 13(1) (January-February 2009)
22. Malik, Z., Bouguettaya, A.: Trust Management for Service-Oriented Environments, 1st edn. Springer, Heidelberg (2009) ISBN:978-1-4419-0309-9
23. Marti, S., Garcia-Molina, H.: Limited Reputation Sharing in P2P Systems. In: Proc. of the 5th ACM Conference on Electronic Commerce, New York, NY, USA, pp. 91–101 (May 2004)
24. Maximillien, E.M., Singh, M.P.: Conceptual Model of Web Service Reputation. SIGMOD Record 31(4), 36–41 (2002)
25. Medjahed, B., Bouguettaya, A.: Customized delivery of e-government web services. IEEE Intelligent Systems 20(6) (November/December 2005)
26. Medjahed, B., Bouguettaya, A., Elmagarmid, A.: Composing Web Services on the Semantic Web. The VLDB Journal 12(4) (November 2003)
27. Niu, J., Chen, Z., Zhang, G.: Towards a subjective trust model with uncertainty for open network. In: Workshops, International Conference on Grid and Cooperative Computing, pp. 102–119 (2006)
28. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford Digital Library Technologies Project (1998)
29. Papazoglou, M.P., Georgakopoulos, D.: Serive-Oriented Computing. Communcications of the ACM 46(10), 25–65 (2003)
30. Resnick, P., Zeckhauser, R.: Trust Among Strangers in Internet Transactions: Empirical Analysis of eBays Reputation System. Advances in Applied Microeconomics, vol. 11. Elsevier Science, Amsterdam (2002)
31. Resnick, P., Zeckhauser, R., Friedman, E., Kuwabara, K.: Reputation Systems. Communication of the ACM 43(12) (December 2000)
32. Rocha, B.G., Almeida, V., Guedes, D.: Increasing qos in selfish overlay networks. IEEE Internet Computing 10(3), 24–31 (2006)
33. Sabater, J., Sierra, C.: Bayesian Network-Based Trust Model. In: Proc. of the first Intl. Joint Conf. on Autonomous Agents and Multiagent Systems, Bologna, Italy, pp. 475–482 (2003)
34. Shibin, Z., Xiang, S., Zhi, Q.: Subjective trust evaluation model based on fuzzy reasoning. International Symposium Electronic Commerce and Security. 1, 328–332 (2009)
35. Sundaresan, N.: Online trust and reputation systems. In: EC 2007: Proceedings of the 8th ACM Conference on Electronic Commerce, pp. 366–367. ACM Press, New York (2007)
36. Tennenholtz, M.: Reputation systems: An axiomatic approach. In: AUAI 2004: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, pp. 544–551. AUAI Press, Arlington (2004)
37. Wang, Y., Vassileva, J.: Trust and reputation model in peer-to-peer networks. In: Proc. of the Third International Conference on Peer-to-Peer Computing, pp. 150–158 (September 2003)
38. Whitby, A., Josang, A., Indulska, J.: Filtering Out Unfair Ratings in Bayesian Reputation Systems. The Icfain Journal of Management Research 4(2), 48–64 (2005)
39. Xiong, L., Liu, L.: PeerTrust: Supporting Reputation-based Trust for Peer-to-Peer Electronic Communities. IEEE Trans. on Knowledge and Data Engineering (TKDE) 16(7), 843–857 (2004)
40. Zhou, R., Hwang, K.: Powertrust: A robust and scalable reputation system for trusted peer-to-peer computing. IEEE Transactions on Parallel and Distributed Systems 18(4), 460–473 (2007)