Artif Intell Rev (2014) 42:835–850
DOI 10.1007/s10462-012-9358-7

Monitoring web services' quality of service: a literature review

Mohd Hilmi Hasan · Jafreezal Jaafar · Mohd Fadzil Hassan

Published online: 30 October 2012

Abstract Monitoring Quality of Service (QoS) compliance is an important procedure in web service environment. It determines whether users' expectations are met, and becomes the vital factor for them to decide whether to continue paying for the service or not. The monitoring is performed by checking the actual services performance against the QoS stated in Service Level Agreement (SLA). In relation to that, the need for monitoring vague QoS specifications in SLA has become more apparent nowadays. This paper reviews the published literature on web services QoS monitoring. A total of 60 selected articles were systematically analyzed. There were 23 of the articles selected through restrictive search criteria while the other 37 were selected based on unrestrictive search criteria. The review shows that little evidence exists on monitoring vague QoS specifications of web services. Providing ability for monitoring QoS that is specified vaguely in SLA could give new insights and implications to web services field. This paper concludes with some recommended future works to construct the theory and perform the empirical research.

Keywords Web service monitoring · Web service QoS · Vague QoS · Fuzzy type-2 · QoS monitoring

1 Introduction

Monitoring for Quality of Service (QoS) compliance has become a vital process in web services environment. The monitoring process is performed by comparing the actual service quality with the expected service quality agreed in contract. Every service comes with a contract between customer and provider to ensure both parties understand the level of service quality that they should respectively expect and provide. In addition, contract also gives benefits to both customers and providers (Liu et al. 2001) in such a way that customers get what they expect for their paid electronic solutions (Angelov and Grefen 2004), while providers

M. H. Hasan (✉) · J. Jaafar · M. F. Hassan
Universiti Teknologi PETRONAS, Tronoh, Perak, Malaysia
e-mail: mhilmi_hasan@petronas.com.my

can prepare effective resources planning and avoid from over-committing the resources to certain services (Allenotor and Thulasiram 2008). This contract is known as Service Level Agreement (SLA) which contains specifications of functional and non-functional QoS. Functional aspect is requirement that describes what tasks or functionalities the service will carry out. On the other hand, non-functional QoS describes how good the service performs the tasks, which leads towards the aspect of quality (Chung and Sampaio do Prado Leite 2009).

Monitoring of web services' QoS is performed by comparing the specified QoS in SLA with the monitored service performance (Yeom et al. 2009). QoS monitoring is important to the extent that it can be used to diagnose the problem that occurs in the service (Leff et al. 2003) and may also become the determining factors for customers to whether continue using the service or not (Zadeh and Seyyedi 2010). The importance of web services' QoS monitoring has lead to numerous studies, as perceived in Beeri et al. (2007), Simmonds et al. (2009), Wu et al. (2008). On the whole, it is evident that QoS monitoring plays important roles in ensuring the quality of web services (Modica et al. 2009). Due to that, QoS monitoring must be carefully planned and structured so that the parties involved will get maximum benefits out of it (Heward et al. 2010). This includes ensuring that the QoS is explicitly specified and its monitoring metrics are defined correctly (Zhou et al. 2004).

The QoS specified in SLA will become the parameters used in the web service monitoring process. Some examples are; "the service availability is at least 98 % in a year", "the response time is not more than 5 seconds" and "the service completes 100 transactions per 60 minutes". These QoS specifications are written in a clear way with exact (certain) expected value for each parameter. Most of the works found in the literature are related to this kind of monitoring of exact parameter values. This paper however, argues that there is a necessity for web services monitoring model to also monitor vague QoS parameters. This allows customers to specify their expected QoS in vague specifications, for example; "the service availability is high", "the response time is low" and "the service has medium completed transactions". It is evident that this type of vague QoS specification leads to the development of an inexperienced-users-centered design of web services. The generalized assumption that customers can formulate their own QoS queries with exact parameter values is therefore argued since they may not even know what the realistic values are (Mobedpour and Chen 2011). Specifying lower than the realistic value will result in customers getting bad quality service. In contrast, specifying higher than the realistic value may result in no matching service is available to be chosen. Furthermore, Rosario et al. (2008) conclude that specifying exact QoS specification is not realistic due to the uncertain nature of web services and network. In addition, even a good QoS may produce inaccurate monitoring results which are caused by the SLA containing vagueness in linguistic definitions (Allenotor and Thulasiram 2008). In general, it is necessary that the web technology has the ability to deal with vague information since this type of information is utilized by many services and applications nowadays (Lukasiewicz and Straccia 2008; Stoilos et al. 2010). This is especially necessary for the web that uses natural language interfaces.

Generally, vagueness is always mistakenly assumed to be similar to uncertainty. Therefore, it is worth to know the difference between them. A simple example to differentiate between the two terms is by looking at the degree of uncertainty and degree of truth of a bottle (Lukasiewicz and Straccia 2008). In relation to the degree of truth, measuring the quantity of water in a statement "the bottle is half full" is determined by the predicate "full". That means the degree of truth that "the bottle is full" reveals the amount of water in that bottle. This is the definition of vagueness. On the other hand, the statement "the probability that the bottle is full is 0.5" does not provide an indication that the bottle is half full. This probability falls under the definition of uncertainty. This paper is interested in covering reviews on the vagueness of QoS specification used in web services monitoring.

The objective of this paper is to present the literature review on web services' QoS monitoring with a purpose to motivate the future works of developing a web services monitoring model for vague non-functional QoS. Section 2 describes the processes involved to collect the related published manuscripts, which are restrictive and unrestrictive search criteria methods. It also contains the critical review on the collected published manuscripts. Section 3 discusses the gaps between the proposed and the published works. This section also presents the significance of the proposed works. Finally, Sect. 4 concludes the paper by summarizing the outcomes of the performed literature review and outlining the future works.

2 Materials and methods

The published literatures reviewed in this paper were collected through two different ways, namely restrictive and unrestrictive search criteria. The subsections that follow contained both types of review.

2.1 Literature review: restrictive search criteria

Journal articles were collected from three sources namely ISI Web of Science, SCOPUS and Google Scholar. The keywords "web service AND monitoring" and "web services AND monitoring" were used when the search was conducted using ISI Web of Science and SCOPUS. On the other hand, the search using Google Scholar used the keyword "web service monitoring". Both SCOPUS and Google Scholar searches were restricted for resources published from 2000 to 2012, while ISI Web of Science was for all years.

The articles yielded from the searches were then refined to ensure that only the related ones were selected for review. In general, the search yielded articles which may be divided into three broad categories, namely performance and QoS monitoring of web services; web services applications used in monitoring systems for business, environment, networks and many other fields; and non-QoS and non-performance monitoring of web services such as monitoring the interactions among services and selection of web services. Only articles from the first category, which is performance and QoS monitoring of web services were selected to be reviewed which yielded a total of 23 journal articles.

This restrictive search criterion provides the advantage of precise search results that are only related to web services' QoS and performance monitoring. In this section, the reviews of all the selected journal articles are presented.

Wu et al. (2010) propose a monitoring for Web Map Service (WMS) to ensure users can select the best available map service during runtime. The problem often arises due to the distributed nature of the WMS services, in which the availability of map services depends on the availability of distributed hosts.

Khaxar and Jalili (2012) present a novel method for runtime monitoring of composite services. The method employs process algebra to express specifications in order to detect engine-related and process-related violations of Business Process Execution Language (BPEL). The monitoring method named as WSCMon performs this performance monitoring by comparing specifications and actual behaviors. One notable feature in this model is the introduction of specifications refinement examination by checking the relation between specifications and the captured behaviors.

Chou et al. (2009) develop a policy-based web services monitoring to improve its QoS. The monitoring system consists of a control center that monitors the status of a number of service monitors. The system is also able to allow users to customize their policies which make the assignment of the monitoring jobs more efficient. Furthermore, the system offers an added benefit to reduce the load of service monitors through predefined policies for CPU and memory.

Porter and Katz (2006) present an approach to develop self adaptive web services by having four design mechanisms, namely statistical techniques to uncover the effects of request in the systems; component monitoring through black-box mechanism, a summarization of statistical information, and techniques to invoke admission control decisions. The black-box monitoring mechanism is proposed due to three main reasons; firstly, to avoid regular monitoring calls that may disrupt the service quality. Secondly, the nature of web services that is rapidly changed and updated makes it hard to perform the monitoring. Thirdly, it may allow distributed component ownership to each site in the environment.

Baazizi et al. (2008) discuss that a complicated task is involved when specifying monitoring expressions of BPEL-specified processes and during extracting their relevant information for monitoring. A monitoring approach that performs an abstraction of BPEL-specified business processes is then proposed. This is done by utilizing business protocols. These abstractions are represented by high level queries which are then compared against services' execution traces contained in the database.

Yeom et al. (2011) design a quality monitoring system for composite web services using a broker. The specifications of service process and policy are done using OWL-S. Any violations are detected by comparing this policy with the observed data, which are collected from the environment using sensor. Meanwhile, Halima et al. (2010) implement a monitoring framework that is part of self healing architecture. In this kind of web services, a complicated configuration is normally needed to associate performance requirements of the applications with the monitoring constraints. The proposed framework is able to provide benefit by reducing this kind of complexity in web services configuration.

Saxena et al. (2009) carry out a research to produce probe based observation in web services' monitoring. The running of web services is observed during its testing and execution. The information such as connection initiation, authentication and service invocation are logged in a file. This file is then used at runtime to perform the monitoring of the services.

Wang et al. (2009) produce a prototype that is able to provide pattern-based service constraints' specifications that correspond to requirements, and perform the monitoring. The monitoring mechanism consists of events that are related to client request, service response, application, resource and management. The constraints' specifications are compared with the collected runtime information in order to perform the monitoring process.

Zadeh and Seyyedi (2011) propose web services monitoring which is based on prediction. The proposed method has the advantages of making the monitoring more intelligent and efficient, while reducing the overhead and cost. The decisions are made base on the predicted conditions which are produced from the current and future values of parameters. In this work, the model is implemented using Neural Network. On the same note, Yu (2007) studies a wrapping-based approach to monitor the performance of web services. The monitoring at client side is very crucial especially in the dynamic and uncertain nature of web services. This study proposes a prediction for future performance of web services based on history data. The data mining technique is implemented to generate the prediction rules.

Yan et al. (2011) discuss on the web service monitoring based on a project that is able to monitor the service status of distributed web servers across the nation. This study proposes the development of an evaluation index system to help in having a comprehensive evaluation method to evaluate the web server's connectivity. Furthermore, this study gives suggestions for managing web services in an optimized ways.

Schaefer (2006) proposes a web services performance monitoring approach based on Application Response Measurement (ARM) standard. By using this approach, web services' messages are attached with metadata that permits them to be customizable and source code independent.

Kalavathy and Seethalakshmi (2009) present a media web service that is composed dynamically, and its performance monitoring using parallel performance monitoring service (PPMS). The PPMS performs the monitoring by examining the dynamic service composition represented in BPEL processes. Among the monitored performance parameters are timeouts, external errors, success rate of each service and fault occurrence. In relation to that, Halima et al. (2008) propose a framework for monitoring and managing self-healing web services that are dynamically composed of services from different providers. The monitoring mechanism is able to analyze the QoS and it observes only SOAP messages passed among services. The framework has an advantage in which it could provide indication of the health status of the web services composition.

Zeng et al. (2007) implement a system that could observe and compute QoS metrics based on service operations. The system contains definitions of IT and business level metrics together with their evaluation formulas. This setup enables the system to retrieve QoS metrics of all service components in the composition, while detecting and routing service operational events systemically.

Liu et al. (2012) present a quality analysis of 2,383 web services and propose a feature of web service quality change. The web service monitoring can be classified into different set of categories based on providers. It is evident that this kind of strategy is suitable for web services monitoring and selection.

Rouached et al. (2005) describe an event-based web services monitoring framework. This framework is associated with a semantic definition of the commitments expressed in the event calculus, instead of specifications that are based on syntactic aspects. It is evident that the framework is able to monitor multi-parties contracts and detect quality violations.

Keller and Ludwig (2003) describe a framework for automated SLA specification and monitoring to ensure the economic and practical benefits to customers. The approach is based on the specification and monitoring by Web Service Level Agreement (WSLA) framework. The framework enables both customers and providers to specify the parameters and the way they are measured. These specifications are then observed when the monitoring system is enforced. Moreover, Asadollah and Thiam (2011) carry out a research to develop a monitoring method for web services response time. The method uses a proxy that is connected to the required services, which is able to calculate the response time.

In addition, a few other works also present web services monitoring mechanisms such as service performance evaluation through constraint based monitoring approach (Zahoor et al. 2009), performance prediction based on historical data and complementary input from other system (Yang et al. 2010) and end-to-end web services performance evaluation using a comprehensive management framework (Guo et al. 2008).

Apparently, all the selected journal articles do not focus on monitoring vague QoS specification for web services SLA as proposed by this paper. Instead, they present different perspectives of web services monitoring such as QoS parameters with exact values. They also present mechanisms and approaches to make web services monitoring more efficient and effective.

2.2 Literature review: unrestrictive search criteria

The search criteria to select the published articles for review were then relaxed in order to gain more related references. A few other keywords were used in the searching processes, such as "monitoring vague QoS", "QoS compliance", "vague SLA" and "fuzzy AND monitoring". In addition, refereed articles from conference proceedings were also chosen for review. Moreover, the chosen articles also include those that were cited by the already-selected articles. However, the publication years for the articles were remained as 2000 until 2012. As a result, 37 more articles were selected in this exercise. This subsection presents the review of these articles. The articles are divided into two categories, namely service monitoring, and service selection and composition.

2.2.1 Web service monitoring

Muller et al. (2009) present the development of a method to produce explanation for the non-compliance in service agreement. The method is based on the WS-Agreement, which is used by the providers to advertise their capabilities through templates. WS-Agreement is also used by customers to define service offer. These two types of specification are then processed by the template publisher to check whether the agreement offer complies with the template or not. In the method, it begins with checking the inconsistencies in the agreement. Next, it checks the compliance between template and agreement offer, and by using constraint satisfaction problem, it produces explanations to describe the causes for the non-compliance. Besides, some other works also present an efficient method for web service monitoring (Raimondi et al. 2007, 2008). The method monitors non-functional QoS based on timeliness constraints such as latency and reliability. QoS in this method is expressed using time automata; hence the violations are detected by checking the occurrence of service by a timed automaton. Overall, it can be concluded that the three articles focus only on monitoring exact QoS specifications in the SLA.

Moreover, Ameller and Franch (2008) introduce a monitoring tool called SALMon which is capable to monitor web service SLA violations. The tool developed in this study focuses on monitoring availability and time behavior characteristics of services. The architecture of SALMon contains three components namely monitors, analyzers and decision makers. The monitors define quality attributes to be measured by the tool. On the other hand, he analyzers perform the task to check the SLA rules, while decision makers carry out the duty to perform corrective actions in order to satisfy the SLA rules. The technique used by the SALMon's analyzers in checking SLA rules violations is however not presented in the paper. Nevertheless, their work only focuses on well-defined QoS specifications using exact values. Another related work, Eze et al. (2010) present e-Health monitoring system. E-Health is a system deployed in SOA-based B2B environment which integrates services from multiple providers. The work extends the conventional SOA process monitoring framework which is designed in policy-based publish/subscribe architecture. The framework has message broker that keeps track all events that have taken place in the e-Health system. It also has a registry of event-based data, which allows its users to observe the past processes or events before making decision whether to utilize any services or not. Thus, web service monitoring in this particular work is only for viewing the processes or events that have taken place to support users' consideration in making decision for service utilization. Quality monitoring is not included in the framework.

Tao et al. (2012) present an approach to predict the QoS trustworthy of web services. The approach is important to be introduced due to the complexity of network as well as the occurrence of illegal operations in web services environment. The model requires the standard UDDI to be extended as a new registry model for web services. The prediction is set up by collecting the QoS data during runtime, which are then processed using an algorithm and clustering method for trustworthy. It is evident that the model can efficiently avoid malicious attack, operation failure and fake information. Besides, Tseng and Wu (2007) implement an expert system that is able to monitor services' web servers. The system is based on knowledge-based management system which retrieves input from experts as part of its result inference process. This management system can detect and resolve stability, reliability and network statuses. These two works are different from the research proposed by this paper in terms of their focus on prediction of QoS trustworthy and diagnosing the possible causes of services' problems. They are not related to performance monitoring of vague QoS specification.

Golshan and Barforoush (2009) develop another type of model, in which the quality of the whole SOA is checked for monitoring purpose. This significantly benefits the users since in most of the cases; a service is composed of multiple subservices. Users need to know the overall quality of the architecture at any time they want, in order to make a decision to choose for the best service. The model they propose is based on combination of Graph Transformation System (GTS) and UML approaches.UML is used to model the quality attributes of the service. It also becomes the input for GTS, which is used to model the static and dynamic parts of an SOA. On the whole, their work focuses on monitoring the overall quality of the architecture. However, there is no consideration on monitoring vague QoS specifications in their work as compared to the research proposed by this paper. In the other related work, Artaiam and Senivongse (2008) model a wider range of service-side quality attributes for monitoring purposes. In this work, they found out that the monitoring of web service is normally limited to a number of quality attributes only namely response time, execution rate and availability. They therefore, extend the model to enable it to monitor more attributes such as accessibility, performance, reliability, security and regulatory. However, there is no specific AI technique being mentioned in their article to explain how the monitoring is performed by the component known as metrics calculating module. On top of that, their work does not focus on monitoring vague QoS of web services.

In a different perspective, Chen et al. (2009) propose a model to evaluate web services' quality attributes for selection decision using Fuzzy C-means artificial immune network memory classifier. The model is based on ontology-oriented evaluation model of semantic web service-based QoS. In their model, there are objective and subjective approaches which are used to decide the weight of the chosen QoS. All of these values of weight are eventually gathered to produce the evaluation judgment. Their work is mainly based on ontology which is not in the scope of the research proposed by this paper. In the other related work, a comprehensive framework is proposed to evaluate SOA governance (Hassanzadeh et al. 2011). This framework gives significant impact to an organization since it is able to portray a clear picture of the success or failure of an SOA. In their framework, three measurement components are indentified namely SOA maturity level, governance maturity level and adoption domains. The framework has been evaluated by 16 experts from the field and it is verified that their framework is more comprehensive than the other previous frameworks. Furthermore, Rosario et al. (2008) present a different perspective in writing a contract by proposing probabilistic QoS. They apply statistical evaluation to monitor QoS compliance. The model assumes that the contract is a distribution on the response time, and compares the observed performance with the stated values in the contract. The model gives warning indication as soon as the first violation is detected in the time distribution. These two works however, do not give any focus on vague QoS as proposed in this paper.

Wetzstein et al. (2009) have extended the web service QoS monitoring to monitoring and analyzing factors that affect business process performance. Business process performance is determined by looking at the achievement of organization's key performance indicators (KPI). Their framework monitors the process events and QoS measurements specified in Web Services Business Process Execution Language (WS-BPEL). It employs machine learning technique to build tree structure that represents dependencies between business processes' KPIs and low level business process metrics and QoS. In this framework, they differentiate the Process Performance Metrics (PPM) monitoring with QoS monitoring. The former relates to metrics that capture single characteristic of business process, while the latter relates to technical metrics such as availability of processor and web service response time. Both types of data give effect to the whole business process performance. Their work however, does not concern on vague QoS specifications as the input for monitoring. The other work, Michlmayr et al. (2010) present a framework known as Vienna Runtime Environment for Service-Oriented Computing (VRESCo) that combines all the entire components in SOA namely service metadata, QoS, service querying, service binding and mediation. The QoS monitoring is performed in their work to evaluate the framework through performance measurements on service querying, binding, mediation and invocation performances. Their work however, focuses only on QoS monitoring of these important processes that exist during service orchestration rather than the performance of service delivery to customers. The research proposed by this paper, on the other hand, focuses on the latter.

2.2.2 Web service selection and composition

Among the popular research topics in web services are service selection and composition. The process of service selection is normally based on QoS specified by customers. The service selection presented in Reiff-Marganiec et al. (2009) is based on non-functional properties and user's context, and is able to not only perform service selection but also evaluate the available services for selection. The model proposed in this work also contains automatic ranking service by considering hard criteria (essential requirements by the user) and soft criteria (essentially preferences). Besides, it is apparent that most of the web services' attributes are uncertain in nature which significantly affects the service selection process. In relation to that, Cheng and Hongbing (2007) propose the uncertainty-aware web services' QoS description and selection model in order to manage this issue. The model is capable to compare QoS stability by utilizing rank sum test of non-parameter test. Their model also contains a decision making model which performs the selection process by considering more than just simple comparison of the mean or variance of parameter values. With this model, the error probability in service selection is reduced since it is not affected by the low and high limits of quality attribute values. Additionally, the QoS-aware service matchmaking during service selection can be made more efficient through users' personalization technique. Personalization is performed by employing data warehousing techniques to capture users' historical service profile (Nagy et al. 2011). By having this, a number of quality enhancements can be gained which includes generating the possible service requests by the tool rather than asking the users to specify them. Moreover, Shin et al. (2009) argue that the current available service selection methods are based on the services' input and output only without providing insights about their actual functional attributes. Thus, service selection and composition may unsatisfactorily provide results that are against users' expectations. In their work, a model that is based on service functional semantics is therefore proposed. The model is implemented using graph model, in which each web service functional specification is searched from the graph during automated service selection and composition. Furthermore, web services composition can also be built based on users' natural language requests (Lim and Lee 2010). In this type of model, each service is published in the registry ontologically by specifying its action and object. Every request written in natural language is processed by the model to produce sentence blocks. These sentence blocks are then further utilized to extract service abstract workflows from the repository. The abstract workflow describes the whole control flow of the composite's constituent services. Using this workflow, actions and objects are then invoked to construct service composition. Furthermore, Wang et al. (2010) present a model that is based on fuzzy for selecting QoS-aware services. The decision making method employed in the model is based on fuzzy liner program. The model provides benefits in terms of its ability to make decision with imprecise and incomplete information and determine the importance weights of QoS criteria. These works are nonetheless not related to QoS monitoring and not involved with vague QoS specification. However, the basic principle of QoS composition and analysis may be considered to be applied in the research proposed by this paper.

In general, web services' QoS monitoring can be performed in two approaches namely server-side and client-side. The former usually gives more accurate monitoring results but it requires entry permission to the actual service implementation which is normally not possible. The client-side monitoring, on the other hand, gives less accurate results since they are affected by other factors that exist throughout the communication. However, client-side monitoring provides independent process that does not require access to the actual service implementation. Due to this trade-offs, Anton Michlmayr et al. (2009) propose a comprehensive web service QoS monitoring and event-based SLA compliance that combines the advantages of both monitoring approaches. Nevertheless, the work does not involve monitoring of vague QoS. In addition, the model proposed by Queiroz et al. (2009) is able to support decision making process in business-driven IT service portfolio management. In making a decision, the model takes into account SLA criteria and the uncertainty nature of business activities such as revenues. The model, however, only deals with exact values of QoS parameters. Moreover, it is obvious that violations of QoS will require the web service system to be reconfigured. The reconfiguration process is normally performed manually which makes it time consuming and costly. The model proposed by Coles et al. (2011) applies stochastic modeling to produce service implementation plan that meets SLA requirements. On top of that, the model also reduces the cost needed through proper automated planning of service implementation. In this model, the duration uncertainty factor is also taken into account in the process of building cost-sensitive service planning. This work, nevertheless, is not related to quality monitoring of web services.

Apparently, a number of researches also give attention to provide mechanism for QoS violation prediction. Zadeh and Seyyedi (2010) argue that continuously monitoring of web services' QoS gives undesirable impact to the whole performance of the service due to the occurrence of overheads every time the monitoring process is invoked. They propose a forecasting model to replace the QoS monitoring in order to reduce this overhead. Their model is based on time series forecasting using Neural Network technique. Moreover, the model proposed by Leitner et al. (2009) is able to predict QoS violations during runtime by using measured and estimated data. The model implements machine learning regression techniques, which are trained using historical processes data. This model is implemented in an architecture that contains several checkpoints, which indicate the points where prediction can be done during the execution of service composition. Apart from that, another two important components in the model are facts and estimates, which are data that is already known and data that is estimated respectively. This model is then extended to PREvent framework which consists of end-to-end composite services evaluation that includes event-based monitoring, prediction of SLA violations and adapting actions to prevent the violations (Leitner et al. 2010). The authors argue that web service prediction as proposed by Zadeh and Seyyedi (2010) cannot totally replace QoS monitoring. Even though overheads do occur but the monitoring plays important role in ensuring that customer's expectations are met and providers' roles are realized. Nevertheless, the prediction of QoS violations is in fact not within the scope of the research proposed by this paper.

3 Discussion

To the authors' knowledge, the works that are closest to the research proposed by this paper are Allenotor and Thulasiram (2008), Mobedpour and Chen (2011), Bacciu et al. (2010), Sherchan et al. (2006). Mobedpour and Chen (2011) discuss that most of QoS-based web service selection methods are not designed for inexperienced users. They argue that the generalized assumption that users may write their own QoS queries is incorrect. Hence, a user-centered design of QoS-based web service selection system is proposed. This system is capable to accept both exact and fuzzy user requirements as the input for the selection process. The fuzzy requirements are handled by clustering the services into three categories namely good, medium and poor. Bacciu et al. (2010) propose service selection model which is based on fuzzy value input. The model performs a matchmaking process between fuzzy-valued QoS parameters specified by users and service capabilities published by providers. The matchmaking process is based on fuzzy similarity measurement. The model is also capable of dynamically updating the service offers by providers through adaptation. The adaptation procedure is made possible through having a continuous runtime monitoring of the services. The monitoring process is based on fuzzy sets, which produces a sequence of performance measurements across time. These series of measurements are compared with the users' QoS through a fuzzy-valued similarity procedure. Moreover, Allenotor and Thulasiram (2008) propose fuzzy grid-QoS framework. They argue that even effective QoS specifications may produce undesirable results due to the vagueness in linguistic definitions of SLA. With regards to grid computing, this kind of issue will result in grid providers overly commit their resources in order to satisfy certain QoS. Hence, the proposed framework is capable of handling vague QoS specifications by users through the implementation of fuzzy logic computation. The framework clusters the linguistic values into five categories namely unacceptable, low, medium, high and outstanding. The evaluation conducted shows that the framework could address the vagueness in linguistic definitions of SLA and hence provides higher grid resources availability by avoiding over committing of computing resources. Furthermore, a fuzzy model is also employed to handle reasoning about web services reputation. This type of reasoning is important to detect any collusions and deceptions in users' ratings towards service providers (Sherchan et al. 2006). In the domain of service reputation, there are always two dimensions occur namely subjective reputation which is based on users' ratings and objective reputation which is based on SLA compliance checking. Hence, the model is proposed by Sherchan et al. (2006), which is capable of taking these two dimensions to produce an effective service reputation system that can avoid the above mentioned issues. In the model, fuzzy technique is used to perform objective reputation by checking for compliance of the actual performance of service providers against QoS in SLA. Then, the results are verified against subjective reputations received from users through fuzzy-based inferring technique.

These four works are different from the research proposed by this paper since they have put the focus on handling vague QoS for service selection process (Mobedpour and Chen 2011;

Bacciu et al. 2010), managing grid resources (Allenotor and Thulasiram 2008) and reasoning reputation (Sherchan et al. 2006). The research proposed by this paper is instead focusing on web services' vague QoS monitoring. Moreover, these four works use fuzzy type-1 technique to resolve the vagueness in linguistic definitions of QoS. In contrast, this paper proposes the use of fuzzy type-2 technique to handle the vagueness specifications. It is stated by Mendel (2003, 2007) that using fuzzy type-1 technique for computing with words is scientifically incorrect. This is due to the fact that fuzzy type-2 can handle more uncertainty and vagueness problems as compared to type-1. Though it is called fuzzy technique, type-1 however, handles uncertainty and vagueness by employing precise membership functions. As a result, the uncertainty and vagueness are simply diminished. This is supported by Dereli et al. (2010), which present three reasons that support the implementation of fuzzy type-2 namely; to handle more uncertainty/vagueness, to investigate and compare fuzzy type-1 and type-2 applications, and to employ fuzzy type-2 in new area of research. Moreover, Castillo and Melin (2002) implement fuzzy type-2 approach for plant monitoring and diagnostic, and perform comparison between type-1 and type2 approaches. The experimental results show that type-2 produces a significant improvement in the monitoring ability. Furthermore, Sepulveda et al. (2007) present the implementation of fuzzy type-2 in designing control systems to minimize the effects of uncertainty produced by the instrumentation elements, environmental noise, and many other factors. This research performs two classes of experiments namely simulations of a feedback control system for a non-linear plant using type-1 and type-2 fuzzy logic controllers, and a non-linear identification problem for time series prediction. Based on the experimental results, the best results are obtained using type-2 fuzzy systems.

Nowadays, fuzzy type-2 has also gained more interest from the researchers. Based on the data retrieved from the two main publication index databases namely ISI Web of Science and Scopus, a significant increment in terms of the number of publications related to fuzzy type-2 can be observed. Fig. 1 presents these numbers for the year 2000 until 2012 for Scopus database. Meanwhile, ISI Web of Science only produces the data for the year 2005 until 2012. It should be noted that a significant increase in the number of publications can be observed for both databases starting with the year 2007. It should also be noted that the number of publications for 2012 is low since the data is taken from only six months of the year. Though fuzzy type-2 was introduced in 1975 (Zadeh 1975), but only recently it started to receive wide acceptance from researchers. One of the main reasons behind this fact is that the previous focus was given on the implementations of fuzzy type-1 in the area of research. After fuzzy type-1 implementation in the research area reaches its maturity level, then only researchers started to explore on the higher level of fuzzy logic namely type-2. The wide acceptance is mainly due to its nature that can handle more vagueness and uncertainty than the fuzzy type-1.

3.1 Significance of the research

Web services have now become essential components of the web technology and are used extensively by users and enterprises (Stavropoulos et al. 2011). The importance of QoS monitoring of web services can mainly be noted from its implementation in Software Oriented Architecture (SOA). Web services are the main components in SOA in which they become the medium that enable the transactions' communications in SOA. The importance of SOA can extensively be perceived after its marriage with the in-trend cloud computing (Jiao et al. 2011; Liang 2011; Patel and Pandey 2010). SOA and cloud computing complement each

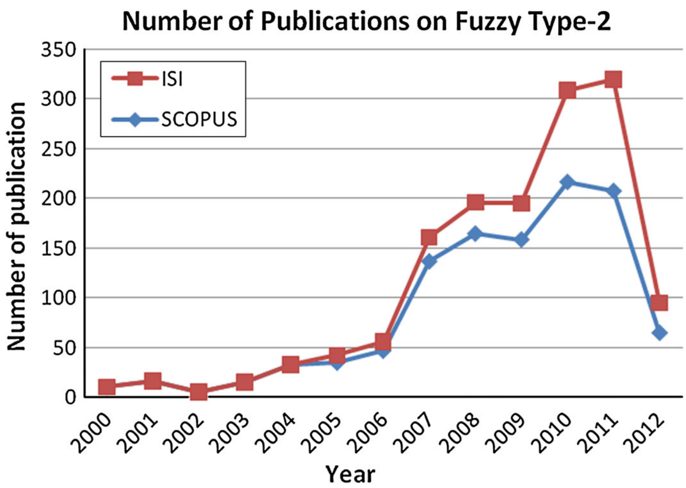

Fig. 1 Number of publications related to fuzzy type-2

other and share the same philosophy of service orientation. Due to its significant advantages such as cost-effective and abstracting the low level implementations, the utilization of cloud computing has been widely used in recent years. The research shows that the worldwide public cloud computing services gained revenue that exceeded \$21.5 *billion in* 2010. *This number is expected to reach* \$72.9 billion in 2015 (International Data Corporation 2012). Another research done by IDC also stated that 14 million jobs worldwide will be produced from the spending on cloud computing services (Microsoft News Center 2012). The implementation of SOA and cloud computing requires the execution of a vital component which is QoS monitoring. The QoS monitoring is a key feature to determine users' satisfaction on SOA and cloud computing services, and to make decision during selection of services. In this paper, the authors propose yet another perspective of QoS monitoring, which applies user-centered approach by allowing QoS to be specified vaguely using natural language.

The importance of vague QoS monitoring is also apparent by means of the implementation of web semantic technology. The web semantic has been growing since its inception, and attracted an increasing users' interest from the fields of biomedical (Hettne et al. 2012), education (Coccoli et al. 2012), knowledge sharing (Wang and Takahashi 2012), information searching (Lee et al. 2012) and many others. Apparently, web semantic technology has to also deal with vague information in its execution (Grouès et al. 2011; Chen et al. 2011; Du and Zhao 2011). The widespread use of vague information in web semantic technology is believed to be a result of users' preferences to use information in natural languages. The authors argue that more of this kind of information will have to be dealt by web semantic technology in the future. Hence, the research on vague QoS monitoring will certainly give contribution to the body of knowledge.

Furthermore, this paper proposes a fuzzy type-2-based solution for vague QoS monitoring. The research proposed by this paper gives opportunity for the exploration of fuzzy type-2 implementation in a new field. Moreover, a comparative study between fuzzy type-1 and type-2 can also be performed which will certainly give contribution to the body of knowledge.

4 Conclusion

Based on the literature presented in this paper, it is evident that the research on vague QoS monitoring of web services is worth to be conducted. The model implies more realistic way to specify web services' QoS. This is especially beneficial for inexperienced users as they can specify the QoS vaguely using natural language.

The proposed model is designed to have capabilities for monitoring single and composite web services' QoS. In a nutshell, fuzzy type-2 is proposed by this paper due to these following three reasons; fuzzy type-2 can handle more vagueness in linguistic definitions of QoS, it is relatively more accurate than fuzzy type-1 as evident in some previous works, and inexistence of type-2 implementation in web services monitoring.

The proposed model is currently undergoing the proof of concept phase. Hence, for future works, a prototype that employs the model will be developed and tested. Furthermore, the research can also be enhanced in terms of having an adaptive metrics measurement values for monitoring.

References

Allenotor D, Thulasiram RK (2008) A fuzzy grid-QoS framework for obtaining higher grid resources availability. In: Proceedings of the 3rd international conference on advances in grid and pervasive computing, pp 128–139

Ameller D, Franch X (2008) Service level agreement monitor (SALMon). Seventh international conference on the composition-based software systems, ICCBSS 2008, pp 224–227

Angelov S, Grefen P (2004) The business case for B2B e-contracting. In: Proceedings of the 6th international conference on electronic commerce, pp 31–40

Artaiam N, Senivongse T (2008) Enhancing service-side QoS monitoring for web services. In: Proceedings of the ninth ACIS international conference on the software engineering, artificial intelligence, networking, and parallel/distributed computing, pp 765–770

Asadollah SA, Thiam KC (2011) Web service response time monitoring: architecture and validation. Theor Math Found Comput Sci Commun Comput Inf Sci 164:276–282

Baazizi MA, Sebahi S, Hacid M-S, Benbernou S, Papazoglou M (2008) Monitoring web services: a database approach. Towards a service-based internet. Lect Notes Comput Sci 5377/2008:98–109

Bacciu D, Buscemi MG, Mkrtchyan L (2010) Adaptive fuzzy-valued service selection. In: Proceedings of the 2010 ACM symposium on applied computing, pp 2467–2471

Beeri C, Eyal A, Milo T, Pilberg A (2007) Monitoring business processes with queries. In: Proceedings of the 33rd international conference on very large data bases, pp 603–614

Castillo O, Melin P (2002) A new hybrid approach for plant monitoring and diagnostics combining type-2 fuzzy logic and fractal theory. In: Proceedings of the 2002 annual meeting of the North American Fuzzy Information Processing Society, NAFIPS, pp 111–116

Chen L, Ha W, Zhang G (2009) A new web service evaluation model with fuzzy C-means artificial immune network memory classifier. In: Proceedings of the 2009 international conference on computational intelligence and security, vol 2, pp 25–29

Chen J, Jia J, Duan L (2011) DOM semantic expansion-based extraction of topical information from web pages. Lect Notes Comput Sci 6988/2011:343–350

Cheng W, Hongbing W (2007) Uncertainty-aware QoS description and selection model for web services. In: Proceedings of the IEEE international conference on services computing, pp 154–161

Chou L-D, Chen C-C, Tseng P-C (2009) Design and implementation of a policy-based monitoring system for web services. J Inf Sci Eng 25(5):1357–1372

Chung L, Sampaio do Prado Leite JC (2009) On non-functional requirements in software engineering. Lect Notes Comput Sci 5600/2009:363–379

Coccoli M, Vercelli G, Vivanet G (2012) Semantic wiki: A collaborative tool for instructional content design. J E-Learn Knowl Soc 8(2):113–122

Coles A, Coles AJ, Clark A, Gilmore S (2011) Cost-sensitive concurrent planning under duration uncertainty for service-level agreements. In: Proceedings of the 21st international conference on automated planning and scheduling, pp 34–41

Dereli T, Baykasoglu A, Altun K, Durmusoglu A, Türksen IB (2010) Industrial applications of type-2 fuzzy sets and systems: a concise review. Comput Ind 62(2):125–137

Du R, Zhao P (2011) Subjective trust evaluation model based on preferences. Commun Comput Inf Sci 215(2):466–472

Eze B, Kuziemsky C, Peyton L, Middleton G, Mouttham A (2010) Policy-based data integration for e-health monitoring processes in a B2B environment: experiences from Canada. J Theor Appl Electron Commer Res 5(1):56–70

Golshan F, Barforoush AA (2009) A new approach for tracing quality attributes in service oriented architecture using graph transformation systems. In: Proceedings of the 14th international computer conference, CSICC 2009, pp 10–16

Grouès V, Naudet Y, Kao O (2011) Combining linguistic values and semantics to represent user preferences. The 6th international workshop on semantic media adaptation and personalization, pp 27–32

Guo N, Gao T, Zhang B (2008) A trusted quality of web services management framework based on six dimensional QoWS model and end-to-end monitoring. Lect Notes Comput Sci 5297:437–440

Halima RB, Guennoun K, Drira K, Jmaiel M (2008) Providing predictive self-healing for web services: a QoS monitoring and analysis-based approach. J Inf Assur Secur 3:175–184

Halima RB, Fki E, Drira K, Jmaiel M (2010) A large-scale monitoring and measurement campaign for Web Services-based applications. Concurr Comput Pract Exp 22(10):1207–1222

Hassanzadeh A, Namdarian L, Elahi S (2011) Developing a framework for evaluating service oriented architecture governance (SOAG). Knowl Based Syst 24(5):716–730. doi:10.1016/j.knosys.2011.02.012

Hettne K, Soiland-Reyes S, Klyne G, Belhajjame K, Gamble M, Bechhofer S, Roos M, Corcho O (2012) Workflow forever: semantic web semantic models and tools for preserving and digitally publishing computational experiments. In: Proceedings of the 4th semantic web applications and tools for the life sciences, pp 36–37

Heward G, Müller I, Jun H, Schneider J-G, Versteeg S (2010) Assessing the performance impact of service monitoring. In: Proceedings of the 21st Australian software engineering conference, IEEE Computer Society, pp 192–201

International Data Corporation (IDC) (2012) IDC cloud research, International Data Corporation. http://www.idc.com/prodserv/idc_cloud.jsp. Accessed 21 June 2012

Jiao Y, Li L, Ye N (2011) Towards a lightweight SOA framework for enterprise cloud computing. The 15th international conference on computer supported cooperative work in design, pp 386–392

Kalavathy GM, Seethalakshmi P (2009) Parallel performance monitoring service for dynamically composed media web services. J Comput Sci 5:487–492

Keller A, Ludwig H (2003) The WSLA framework: specifying and monitoring service level agreements for web services. J Netw Syst Manag 11(1):57–81

Khaxar M, Jalili S (2012) WSCMon: runtime monitoring of web service ochestration based on refinement checking. Serv Oriented Comput Appl 6(1):1–17

Lee M, Kim W, Park S (2012) Searching and rankinxg method of relevant resources by user intention on the semantic web. Expert Syst Appl Int J 39(4):4111–4121

Leff A, Rayfield JT, Dias DM (2003) Service-level agreements and commercial grids. IEEE Internet Comput 7(4):44–50

Leitner P, Wetzstein B, Rosenberg F, Michlmayr A, Dustdar S, Leymann F (2009) Runtime prediction of service level agreement violations for composite services. In: Proceedings of the 2009 international conference on service-oriented computing, pp 176–186

Leitner P, Michlmayr A, Rosenberg F, Dustdar S (2010) Monitoring, prediction and prevention of SLA violations in composite services. In: Proceedings of the 2010 IEEE international conference on web services, pp 369–376

Liang S. (2011) The design and realization of cloud computing framework model based on SOA. Adv Mater Res 171(172):696–701

Lim J, Lee KH (2010) Constructing composite web services from natural language requests. Web Semant 8(1):1–13. doi:10.1016/j.websem.2009.09.007

Liu Z, Squillante MS, Wolf JL (2001) On maximizing service-level-agreement profits. In: Proceedings of the 3rd ACM conference on electronic commerce, pp 213–223

Liu G-Q, Zhu Z-L, Liu Y (2012) Quality analysis and monitoring strategy research for web service in the internet. Technol Educ Learn Adv Intell Soft Comput 136:541–548

Lukasiewicz T, Straccia U (2008) Managing uncertainty and vagueness in description logics for the semantic web. Web Semant 6(4):291–308

Mendel JM (2003) Fuzzy sets for words: a new beginning. In: Proceedings of the 12th IEEE international conference on fuzzy systems, vol 1, pp 37–42

Mendel JM (2007) Computing with words: Zadeh, Turing, Popper and Occam. IEEE Comput Intell Mag 2(4):10–17

Michlmayr A, Rosenberg F, Leitner P, Dustdar S (2009) Comprehensive QoS monitoring of web services and event-based SLA violation detection. In: Proceedings of the 4th international workshop on middleware for service oriented computing, pp 1–6

Michlmayr A, Rosenberg F, Leitner P, Dustdar S (2010) End-to-end support for QoS-aware service selection, binding, and mediation in VRESCo. IEEE Trans Serv Comput 3(3):193–205

Microsoft News Center (2012) Cloud computing to create 14 million new jobs by 2015, Microsoft. http://www.microsoft.com/en-us/news/features/2012/mar12/03-05CloudComputingJobs.aspx. Accessed 20 June 2012

Mobedpour D, Chen D (2011) User-centered design of a QoS-based web service selection system. Serv Oriented Comput Appl 1–11. doi:10.1007/s11761-011-0091-x

Modica GD, Tomarchio O, Vita L (2009) Dynamic SLAs management in service oriented environments. J Syst Softw 82(5):759–771

Muller C, Resinas M, Ruiz-Cortes A (2009) Explaining the non-compliance between templates and agreement offers in WS-Agreement. In: Proceedings of the 7th international joint conference on service-oriented computing, pp 237–252

Nagy W, Mokhtar HMO, El-Bastawissy A (2011) A flexible tool for web service selection in service oriented architecture. Int J Adv Comput Sci Appl 2(12):191–201

Patel SV, Pandey K (2010) Design of SOA based framework for collaborative cloud computing in wireless sensor networks. Int J Grid High Perform Comput 2(3):60–73

Porter G, Katz RH (2006) Effective web service load balancing through statistical monitoring. Commun ACM Self Manag Syst 49(3):48–54

Queiroz M, Moura A, Sauve J, Bartolini C, Hickey M (2009) A model for decision support in business-driven IT service portfolio management using SLA-dependent criteria and under uncertainty. In: Proceedings of the international conference on management of emergent digital ecosystems. doi:10.1145/1643823.1643874

Raimondi F, Skene J, Emmerich W, Wozna B (2007) A methodology for online monitoring non-functional specification of web-services. The first international workshop on property verification for software components and services (PROVECS'07), pp 50–59

Raimondi F, Skene J, Emmerich W (2008) Efficient online monitoring of web-service SLAs. In: Proceedings of the 16th ACM SIGSOFT international symposium on foundations of software engineering, pp 170–180

Reiff-Marganiec S, Yu HQ, Tilly M (2009) Service selection based on non-functional properties. Service-oriented computing—ICSOC 2007 workshops, pp 128–138

Rosario S, Benveniste A, Haar S, Jard C (2008) Probabilistic QoS and soft contracts for transaction-based web services orchestrations. IEEE Trans Serv Comput 1(4):187–200

Rouached M, Perrin O, Godart C (2005) A contract-based approach for monitoring collaborative web services using commitments in the event calculus. Web Inf Syst Eng Lect Notes Comput Sci 3806/2005:426–434

Saxena N, Goel A, Singh D (2009) A probe-based observability mechanism for monitoring of web services. Int J Recent Trends Eng 1(1):600–602

Schaefer J (2006) An approach for fine-grained web service performance monitoring. Distrib Appl Interoperable Syst Lect Notes Comput Sci 4025/2006:169–180

Sepulveda R, Castillo O, Melin P, Rodriguez-Diaz A, Montiel O (2007) Experimental study of intelligent controllers under uncertainty using type-1 and type-2 fuzzy logic. Inf Sci Int J 177(10):2023–2048. doi:10.1016/j.ins.2006.10.004

Sherchan W, Loke SW, Krishnaswamy S (2006) A fuzzy model for reasoning about reputation in web services. In: Proceedings of the 2006 ACM symposium on applied computing, pp 1886–1892

Shin D-H, Lee K-H, Suda T (2009) Automated generation of composite web services based on functional semantics. Web Semant Sci Serv Agents World Wide Web 7(4): 332–343. doi:10.1016/j.websem.2009.05.001

Simmonds J, Gan Y, Chechik M, Nejati S, O'Farrell B, Litani E, Waterhouse J (2009) Runtime monitoring of web service conversations. IEEE Trans Serv Comput 2(3):223–244

Stavropoulos TG, Vrakas D, Vlahavas I (2011) A survey of service composition in ambient intelligence environments. Artif Intell Rev 1–24. doi:10.1007/s10462-011-9283-1

Stoilos G, Stamou G, Pan JZ (2010) Fuzzy extensions of OWL: logical properties and reduction to fuzzy description logics. Int J Approx Reason 51(6):656–679

Tao Q, Chang H-Y, Gu C-Q, Yi Y (2012) A novel prediction approach for trustworthy QoS of web services. Expert Syst Appl Int J 39(3):3676–3681. doi:10.1016/j.eswa.2011.09.060

Tseng JCR, Wu C-H (2007) An expert system approach to improving stability and reliability of web service. Expert Syst Appl Int J 33(2):379–388. doi:10.1016/j.eswa.2006.05.024
Wang K, Takahashi A (2012) Semantic web based innovative design knowledge modeling for collaborative design. Expert Syst Appl Int J 39(5):5616–5624
Wang Q, Shao J, Deng F, Liu Y, Li M, Han J, Mei H (2009) An online monitoring approach for web service requirements. IEEE Trans Serv Comput 2(4):338–351
Wang P, Chao K-M, Lo C-C (2010) On optimal decision for QoS-aware composite service selection. Expert Syst Appl Int J 37(1):440–449. doi:10.1016/j.eswa.2009.05.070
Wetzstein B, Leitner P, Rosenberg F, Brandic I, Dustdar S, Leymann F (2009) Monitoring and analyzing influential factors of business process performance. In: Proceedings of the 2009 IEEE international enterprise distributed object computing conference (EDOC 2009), pp 141–150
Wu G, Wei J, Huang T (2008) Flexible pattern monitoring for WS-BPEL through stateful aspect extension. In: Proceedings of the 2008 IEEE international conference on web services, pp 577–584
Wu H, Li Z, Zhang H, Yang C, Shen S (2010) Monitoring and evaluating the quality of web map service resources for optimizing map composition over the internet to support decision making. Comput Geosci 37(4):485–494
Yan S, Shang J, Liu L (2011) Design and implementation of status monitoring system for e-learning web service. Commun Comput Inf Sci 233:158–165
Yang Y, Luo D, Zhang C (2010) A multiple system performance monitoring model for web services. Agents Data Min Interact Lect Notes Comput Sci 5980/2010:149–161
Yeom G, Tsai W-T, Bai X, Lee Y (2009) Design of a contract-based web services QoS management system. The 29th IEEE international conference on distributed computing systems workshops, pp 306–311
Yeom G, Tsai W-T, Bai X, Lee Y (2011) A design of policy-based composite web services QoS monitoring system. Int J Crit Comput Based Syst 2(1):79–91
Yu L (2007) Applying software wrapping on performance monitoring of web services. INFOCOMP J Comput Sci 6(3):1–6
Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning—1. Inf Sci 8:199–249
Zadeh MH, Seyyedi MA (2010) Qos monitoring for web services by time series forecasting. 2010 3rd IEEE international conference on computer science and information technology (ICCSIT), pp 659–663
Zadeh MH, Seyyedi MA (2011) Applying a predictive approach for QoS monitoring in web service. Commun Comput Inf Sci 142(1):237–241
Zahoor E, Perrin O, Godart C (2009) An integrated declarative approach to web services composition and monitoring. Lect Notes Comput Sci 5802:247–260
Zeng L, Lei H, Chang H (2007) Monitoring the QoS for web services. Serv Oriented Comput Lect Notes Comput Sci 4749/2007:132–144
Zhou C, Chia L-T, Lee BS (2004) DAML-QoS ontology for web services. In: Proceedings of the IEEE international conference on web services, 2004, pp 472–479