Credible Service Selection in Cloud Environments

by

Lie Qu

A thesis submitted in fulfilment of

the requirements for the award of the degree

Doctor of Philosophy

from

Department of Computing

Faculty of Science

MACQUARIE UNIVERSITY

Supervisors: A/Prof. Yan Wang

Prof. Mehmet A. Orgun

2016

Statement of Candidate

I certify that the work in this thesis entitled **"Credible Service Selection in Cloud Environments"** has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

Lie Qu

Lie Qu

3 July 2016

To my parents,

who are the pillars of my life.

In memory of my cousin,

Hezheng Zhao.

Abstract

With the development of the Internet, cloud computing has become the most popular paradigm for on-demand provision of computing resources and service solutions. Due to the high flexibility of cloud computing, a myriad of cloud-based services are designed and implemented to satisfy consumers' diverse needs, and thus new challenges have arisen in cloud service selection. One big challenge is how to select the most suitable cloud service for potential consumers according to their customized requirements. Another big challenge is how to perform cloud service selection with high effectiveness and accuracy, i.e., identifying unreasonable assessments and eliciting credible assessments. This thesis aims to systematically investigate the above two challenges of cloud service selection. The main contributions of this thesis are summarized as follows:

- In prior studies, cloud service selection usually depends on quantitative performance analysis without considering cloud consumers' opinions on service performance. This causes a problem that some vital performance aspects, which can hardly be evaluated through objective monitoring and testing, e.g., data privacy and after-sales services, are ignored in cloud service selection. To solve this problem, we propose a novel model of cloud service selection by aggregating both subjective assessments from ordinary cloud consumers and objective assessments extracted through quantitative analysis. By comparing and aggregating such assessments, the result of service selection can reflect the overall quality of a cloud service with less bias caused by unreasonable assessments.

- We further consider the contexts of cloud assessments and cloud service requesters in our proposed model. In this model, a cloud consumer is allowed to specify under what condition (e.g., specify a particular location or a particular

period of time) he/she would like to consume a cloud service. Then the service selection is carried out based on the consumer's context. In this way, our cloud service selection model can more effectively reflect potential cloud consumers' customized requirements.

- In order to improve the accuracy of cloud service selection, we propose a novel approach to evaluate the credibility of cloud assessments. Considering the dynamic nature of cloud services, the proposed approach is based on the continual assessments over time, and thus has the ability to not only evaluate the dynamic performance of cloud services, but also resist user collusion of providing malicious assessments. Through this approach, more credible assessments are selected as input for further cloud service selection.

- In addition to the assessment credibility evaluation, we have found another way to further improve the accuracy of cloud service selection. We propose a novel incentive mechanism, through which a cloud user would have sufficient incentives to provide continual and truthful assessments of cloud services. This would benefit both the dynamic evaluation of cloud performance and the further cloud service selection. Furthermore, the proposed incentive mechanism allows users to provide uncertain assessments when they are not sure about the real performance of cloud services, rather than providing untruthful or arbitrary assessments which may greatly affect the accuracy of service selection.

All the models, approaches and mechanism proposed in this thesis have been validated and evaluated through sufficient experiments and theoretical analysis. The results have demonstrated that the proposed approaches and mechanisms outperform the existing work of cloud service selection.

Acknowledgments

The thesis would not have been possible without the help of many people who contributed to my research in one way or another.

First of all, I would like to express my sincere thanks to my supervisors, A/Prof. Yan Wang and Prof. Mehmet A. Orgun. I am very grateful for their kind and patient supervision. They have led me in the correct direction over the past few years not only with their knowledge and experience, but also with the attitude for high quality research.

I would also appreciate all the co-authors of my publications, Prof. Ling Liu, Prof. Athman Bouguettaya, Prof. Huan Liu, and Prof. Duncan S. Wong. My academic journey would not have been so rewarding without their kindness and wisdom.

I wish to express my thanks to my colleagues, Dr. Guanfeng Liu, Dr. Haibin Zhang, Jun Zou, Dr. Xiaoming Zheng and Ye bin, who provided very helpful suggestions for my research.

I would like to thank the administrative staff of Department of Computing, Donna Hua, Melina Chan, Sylvian Chow, Camille Hoffman, Jackie Walsh and Fiona Yan for their warm support and kind help.

Most importantly, I would like to thank my parents, Zhaotai Qu and Xu Zhao. Their love, support and encouragement are the most valuable treasures in my life.

Publications

This thesis is based on the research work I have performed with the help of my supervisors and other colleagues during my PhD program at the Department of Computing, Macquarie University between 2012 and 2016. Some parts of my research have been published/submitted in the following papers:

[1] **Lie Qu**, Yan Wang, Mehmet A. Orgun: An Uncertain Assessment Compatible Incentive Mechanism for Eliciting Continual and Truthful Assessments of Cloud Services, submitted to the 23rd IEEE International Conference on Web Services, ICWS 2016 (**CORE2014**[1] **Rank A**).

[2] **Lie Qu**, Yan Wang, Mehmet A. Orgun: A Novel Incentive Mechanism for Truthful Performance Assessments of Cloud Services, accepted by International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2016 (**CORE2014 Rank A***).

[3] **Lie Qu**, Yan Wang, Mehmet A. Orgun, Ling Liu, Huan Liu, Athman Bouguettaya: CCCloud: Context-Aware and Credible Cloud Service Selection based on Subjective Assessment and Objective Assessment. IEEE Transactions on Services Computing (TSC) 8(3): 369-383 (2015) (**Impact Factor: 3.049**).

[4] **Lie Qu**, Yan Wang, Mehmet A. Orgun, Ling Liu, Athman Bouguettaya: Context-Aware Cloud Service Selection based on Comparison and Aggregation of User Subjective Assessment and Objective Performance Assessment. 21th IEEE International Conference on Web Services (ICWS 2014): 81-88 (**Research Track, CORE2014 Rank A**).

[1] CORE refers to the Computing Research and Education Association of Australasia (www.core.edu.au).

[5] **Lie Qu**, Yan Wang, Mehmet A. Orgun, Ling Liu, Athman Bouguettaya: Cloud Service Selection based on Contextual Subjective Assessment and Objective Assessment. 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014): 1483-1484 (**CORE2014 Rank A***).

[6] **Lie Qu**, Yan Wang, Mehmet A. Orgun, Duncan S. Wong, Athman Bouguettaya: Evaluating Cloud Users' Credibility of Providing Subjective Assessment or Objective Assessment for Cloud Services. 12th International Conference on Service Oriented Computing (ICSOC 2014): 453-461 (**CORE2014 Rank A**).

[7] **Lie Qu**, Yan Wang, Mehmet A. Orgun: Cloud Service Selection based on the Aggregation of User Feedback and Quantitative Performance Assessment. 10th International Conference on Services Computing (SCC 2013): 152-159 (**Research Track, CORE2014 Rank A**).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As the most effective and efficient paradigm for on-demand computing resource sharing and flexible service provision in recent years, cloud computing has been attracting huge attention. According to the definition from the National Institute of Standards and Technology (NIST), cloud computing can be defined as follows [109]:

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Although the term "cloud computing" has become popular only since several years ago, it is not a totally new technology. It is the evolutionary result of many existing technologies and methodologies. Hence, cloud computing has some of the same characteristics with other computing concepts. For example, from the perspective of computing resource delivery, cloud computing is similar to grid computing [13]. The consumption of the computing resources under both concepts can be treated similarly as electricity from the real power grid. In addition, cloud computing is also similar to the concept of utility computing from the resource pricing perspective [11]. All the resources under both concepts can be metered, so that consumers only need to pay for the resources they use. However, the concept of cloud computing is still different from either grid computing or utility computing. For example, distributed computing is the backbone of grid computing, but cloud computing can also support non-distributed infrastructure. On the other hand, utility computing mainly focuses on the pricing model of the metered resource delivery, but cloud computing not only focuses on the resource pricing model but also resource deployment, application, dynamic sharing, etc. The essential characteristics of cloud computing are *on-demand self-service, broad network access, resource pooling, rapid elasticity* and *measured service* [109].

In recent years, more and more individuals and organizations have been moving their daily work into cloud environments, because of its features of flexibility and low cost. According to consumers' diverse requirements, a variety of cloud-based services can be designed and implemented, and some even can be customized for consumers' needs. Moreover, cloud consumers are also allowed to develop their own cloud services based on existing cloud infrastructures or platforms, and provide these services to other consumers. In addition to the high flexibility of cloud computing, its *pay-as-you-go* [11] model hugely reduces the cost of accessing computing resources and services for cloud consumers. Compared to traditional computing service patterns, cloud consumers do not need to worry about large upfront cost on expensive computing hardware and software before setting up their work. Instead, they can access a shared computing resource or service pool, and pay while using the pool according to their different needs.

Due to these advantages of cloud computing, diverse cloud services are designed and implemented. Fig. 1.1 [2] illustrates some popular cloud service providers. With the emergence of various cloud services, many new challenges have arisen for cloud consumers. One main challenge is how to discovery and select the most suitable one from a myriad of alternative cloud services. Web service selection in general service-oriented environments has been sufficiently studied in the literature, e.g., [148, 101, 16, 55, 156, 183, 179, 180, 108, 107, 165, 163, 96], but service selection in cloud environments has not been paid close attention with the widespread use of cloud computing until recently. Although cloud services have a very close relation to web services, the research focus of cloud service selection is not totally identical to that of traditional web service selection due to the features of cloud services. Some techniques and approaches applied for ordinary web service selection may not be ap-

Figure 1.1: Cloud Service Providers

plicable in cloud environments, or need to be improved or extended when considering the cloud features.

1.1 Challenges in Cloud Service Selection

Although cloud services are very similar to traditional web services from the perspective of ordinary consumers, e.g., SaaS clouds can be nearly considered as web-based applications, and PaaS clouds and IaaS clouds need to be accessed from web-based interfaces, new challenges in cloud service selection have arisen due to the features of cloud computing, and should be addressed for large numbers of potential cloud consumers. Compared to ordinary web services selection, cloud service selection has the following main challenges:

- **More types of services:** everything in cloud environments can be taken as services, e.g., Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and

Software-as-a-Service (SaaS), and any individual or organization can develop their own cloud services based on existing cloud infrastructure or platforms. Hence, there are a huge number of various cloud services with different functionalities or non-functional performance. For diverse types of cloud services, consumers' requirements may be quite different. Hence, users' requirements should be paid more attention in cloud service selection, e.g., identifying necessary requirements according to cloud types, and dealing with the subjective vagueness of requirements.

- **More performance criteria:** traditional web service selection usually considers only a few crucial service performance attributes, e.g., service availability, response time, throughput, etc. Cloud computing can provide the services with high complex functionalities, e.g., IaaS or PaaS clouds. Hence, many unique cloud performance attributes, e.g. scalability, elasticity, reliability, payment policies, security, etc., should be taken into account when carrying out cloud service selection. Moreover, some specific performance aspects need to be considered particularly according to different requirements, e.g., *persistent storage* and *service networking* are discussed in [85] for selecting public clouds. Hence, identifying crucial performance attributes is very important for cloud service selection.

- **More difficult performance evaluation:** due to the diversity of performance criteria of cloud services, accurately collecting and evaluating these criteria becomes much more difficult. In general, the values of the criteria are collected from pre-defined service level agreement (SLA) [174] or service performance monitoring and testing carried out by cloud providers or end-users. It should be noted that SLA cannot reflect the real performance of cloud services, and service evaluation at the runtime requires expertise and extra costs of consuming cloud services, thus service evaluation is usually provided by cloud providers. However, such an evaluation cannot be fully trusted. That is because, providers have direct interest in service performance, and service-side performance evaluation may not accurately reflect the experiences of end-users.

Prior to cloud service selection, an evaluation of cloud services should be performed first from the perspective of end-users. In the literature, there are two types of approaches which can be used to conduct such an evaluation. The first type of approaches is based on objective performance assessment from ordinary QoS parameters (Quality-of-Service, e.g., service response time, availability and throughput) [194, 21, 161] and predesigned benchmark testing [85, 14, 84]. As cloud services are highly virtualized, some methods and tools for traditional IT computation measurements can be appropriately applied in cloud environments. By combining these methods and tools according to the characteristics of cloud services, many metrics can be quantitatively assessed (e.g., the speed of CPU and I/O). The second type of approaches is based on ordinary consumers' subjective assessments extracted from their subjective feedback for every concerned aspect of cloud services [137, 118]. In this type of approaches, cloud services are usually treated like traditional web services, thus some rating-based reputation systems [148, 87, 101] can be utilized for cloud service selection.

Nevertheless, these two types of cloud service evaluation approaches have their own limitations. That is because, firstly, objective performance assessment can only be carried out for the performance aspects which can be easily quantified. Conversely, objective assessments are not appropriate for those aspects which are quite hard to quantify, such as data security, privacy and after-sales services. On the other hand, subjective assessments have the risk of inaccuracy since users' subjective feelings are very likely to contain bias and not to reflect the real situations of cloud performance. Furthermore, there may even be malicious consumers who give unreasonable subjective assessments with the intention to deceive others and/or benefit themselves in some cases [117, 101]. As a result, the accuracy of overall subjective assessments for cloud services can be significantly affected. Due to all the reasons mentioned above, credible cloud service evaluation from end-users' perspective should be a main research topic in cloud service selection.

- **More dynamic performance:** cloud services are usually consumed simultaneously by a large number of users. The features of cloud computing *scalability* and *elasticity* guarantee that every user can experience stable service performance at the same time, even the peak-load time, and cloud services have the capability to distribute required resources to new users near real-time [102]. However, in the real-world situations, cloud services may have unstable performance due to the constantly changing number of users. Hence, the ability of cloud providers dynamically provisioning computing resources should be evaluated in multi-tenancy environments [81], which has not been adequately studied in traditional web service selection.

  In addition, the performance of cloud service could vary considerably over time due to the dynamic computing resource allocation and the frequent variation of the number of cloud consumers. Hence, there should be a way to evaluate the dynamic performance of cloud services over time. To this end, cloud consumers should be motived to provide the assessments of cloud service regularly from user-end over time. However, consumers usually do not have sufficient incentives to do so, and may even provide biased or arbitrary assessments, which may greatly affect the efficiency of cloud service selection. To solve this problem, an incentive mechanism for eliciting consumers' continual and truthful assessments should be designed, and the credibility of cloud assessments should be carefully evaluated before service selection.

- **More varied requirements:** in addition to hardly obtain comprehensive and accurate performance evaluation of cloud services, cloud consumers' requirements vary considerably. Because of the advantages of cloud services, more and more organizations and individuals move their work into clouds. The requirements from different consumers are quite different. For example, Some industrial consumers may give priority to the overall performance of cloud services as well as some particular performance aspects, e.g., data privacy and service scalability. And individual consumers may take an important consideration on the cost of services. For different types of consumers, an effective selection approach should be compatible with various consumers' customized needs and conditions. In some cases, consumers may not explicitly know what he/she needs for cloud services. Thus, the uncertainty of consumers' requirements should be addressed in an effective service selection approach.

1.2 Motivation

In this section, an example is introduced to motivate our work in practice. Considering a health center processing a large amount of sensitive customer data every day, the security and privacy of customer data have a crucial impact on the center's reputation. If the center plans to move its work into cloud environments in order to reduce daily costs, a suitable cloud provider which has a very good reputation on data security and privacy needs to be selected. In addition, as the health center is not a professional IT organization, comprehensive and high quality after-sales services are highly desired. Moreover, a variety of encryption approaches need to be frequently applied due to the sensitivity of customer data. Hence, the speed of data encryption and decryption is a big concern for the center.

In this example, in addition to the ordinary objective performance attributes of cloud services (e.g., availability and response time), which can be quantitatively monitored, the speed of data encryption and decryption is a big concern for the center since a variety of encryption approaches need to be frequently applied due to the sensitivity of customer data. Such performance attributes may need to be specifically tested through benchmark testing. Moreover, the health center needs to carefully consider the data privacy and quality of after-sales services, which are very hard to quantify, but may easily be assessed through subjective assessments provided by other cloud consumers who consume the same cloud services. Hence, a service selection approach considering both subjective and objective assessments should be highly desired by the center. In addition, before carrying out service selection, the credibility of cloud users on how truthfully they provide subjective or objective assessments should be evaluated first to guarantee service selection accuracy.

On the other hand, objective assessments of cloud services are usually provided by professional monitoring and testing organizations, such as CloudHarmony[1] and CloudSleuth[2]. Such organization usually have a number of cloud monitoring and testing centers all over the world. Thus, there may be multiple testing parties in one city or region. In addition, ordinary cloud consumers who provide subjective assessments of cloud services may also be located all around the world. That means that all subjective or objective assessments are provided under their own contexts (e.g., different *locations* and *time*). Therefore, it should be considered that, assessment contexts can significantly influence the result of assessments. According to the objective statistics from CloudSleuth, the *response time* of a cloud service varies significantly under different worldwide QoS monitoring and testing centers, and generally increases with the increasing distances between the cloud provider and these centers because of the increasing length of the network routes of cloud service delivery. Meanwhile, the assessed results of response time can also be affected by the time of a day, in other words, how busy the cloud service and the network accessed by the centers can be at different times of a day. Therefore, both objective assessments and subjective assessments can be affected by different assessment contexts. When the health center asks for cloud service selection, it needs to determine which assessments should be trusted more according to its own context, and then be aware of how truthfully these assessments are provided (i.e., cloud users' credibility). The assessment evaluation should be a crucial aspect considered in service selection.

[1] www.cloudharmony.com
[2] www.cloudsleuth.net

Furthermore, in order to maintain the stability of the services offered by the center, the long-term performance of the selected cloud service is also an important consideration for the center. Once the performance of the cloud service becomes rotten, the center would hope to find another alternative service as soon as possible. To this end, the continual assessments of the service play a very vital role in service selection. However, at the current stage, wherever in literature or in practice, the continual assessments have not been paid enough attention in service selection, and there is no existing mechanism proposed for motivating cloud users to provide assessments regularly over time. Thus, evaluating the dynamic performance of cloud services is still a challenging problem. This flaw would lead the health center lack of confidence to consume cloud service.

1.3 Contributions

In a nutshell, the core contribution of the thesis is to enhance the effectiveness of cloud service selection through comprehensive assessments. In order to address the significant and challenging issues introduced above, this thesis makes contributions in the following four major aspects.

1. The first contribution of the thesis focuses on cloud service selection based on comprehensive assessments.

   (a) Different from all the existing approaches, we propose a novel model of cloud service selection based on the comparison and aggregation of both subjective assessments from ordinary cloud consumers and the objective assessments through quantitative performance monitoring or testing. A framework is proposed to support the proposed model.

   (b) In our model, cloud performance is comprehensively evaluated through the aggregated results of all the subjective assessments and the objective assessments through a fuzzy simple additive weighting system [26]. Furthermore, we consider the situation where ordinary cloud consumers' subjective assessments may be biased and inaccurate since they are usually not professional IT staff or even some of them are malicious. Thus, in our model, ordinary consumers' subjective assessments are compared to the objective assessments for the same cloud services, so that unreasonable subjective assessments can be filtered out before aggregation. That is because objective assessments do not contain subjective bias, and are thus more reliable than subjective assessments.

2. The second contribution of the thesis focuses on cloud service selection considering contextual assessments.

   (a) We extend our proposed model of cloud service selection by taking the contexts of assessments into account. In the extended version, according to a potential cloud consumer's requirements, the objective assessments are first applied as a benchmark to filter out biased or unreasonable subjective assessments. In order to guarantee the accuracy of such filtering, our model considers two assessment features (i.e., *location* and *time*) in contexts, which can commonly affect both objective assessments and subjective assessments.

   (b) The process of the filtering is based on the context similarity between objective assessments and subjective assessments, i.e., the more similar the context, the more reliable subjective assessments, so that the benchmark level can be dynamically adjusted according to the corresponding context similarity. We propose a novel approach for computing the similarity between assessment contexts based on the bipartite SimRank algorithm [69] After such filtering, the final aggregated results can reflect the overall performance of cloud services according to potential users' personalized requirements and context.

3. The third contribution of the thesis focuses on evaluating cloud users' credibility of providing subjective assessments or objective assessments for cloud services.

   (a) We propose a model for evaluating cloud users' credibility of providing subjective assessments or objective assessments, where subjective assessments are from ordinary cloud consumers (called Ordinary Consumers, *OC* for short), and objective assessments are from professional cloud performance monitoring and testing parties (called Testing Parties, *TP* for short). The credibility of OCs and TPs is respectively represented by *trustworthiness* of OCs and *reputations* of TPs, which are separately evaluated and can be influenced by each other. For this reason, our model can resist collusion among cloud users providing untruthful assessments to some extent. Through our model, a successful collusion attack would become very difficult in practice since a large number of cloud users would have to be involved in such collusion.

   (b) We propose an improved model of cloud service selection considering both contextual assessments and cloud users' credibility. When a potential cloud consumer requests cloud service selection, the context similarities between the consumer and different *TP*s are first computed to determine which *TP*(s) is/are more reliable. Then, all the *TP*s are grouped according to their contexts. Cloud service selection is carried out independently in every context group of *TP*s. In each context group, when the benchmark filtering is carried out, the context similarities between subjective assessments and objective assessments are computed to determine the benchmark levels, so that dynamic benchmark levels can be set to make such benchmark filtering more accurate.

       In the meantime, an *OC* in a context group is considered more credible if his/her historical subjective assessments are more similar with the majority of subjective or objective assessments from *OC*s or *TP*s. In addition, the credibility of an *OC* can also be affected by the difference of variation trends between the *OC*'s subjective assessments and *TP*s' objective assessments over time, so that the dynamic performance of cloud services can be evaluated based on the variation trends of assessments. On the other hand, the credibility of a *TP* depends on the difference between its objective assessments and the majority of objective or subjective assessments from *TP*s or *OC*s. That makes our model able to resist user collusion. Through the proposed model, the selection results can comprehensively and accurately reflect various cloud consumers' needs in cloud service selection.

4. The fourth contribution of the thesis focuses on the design of an uncertain-assessment-compatible incentive mechanism for eliciting continual and truthful assessments of cloud services.

   (a) In order to achieve high accuracy in cloud service selection, the contextual assessments and cloud users' credibility are studied in our proposed models. Such a way can be considered as a passive way of improving selection accuracy. On the other hand, through a suitable incentive mechanism in cloud environments, cloud users can be motivated to always provide truthful assessments, which can be considered as an active way of improving selection accuracy. To this end, we design a novel incentive mechanism based on Game Theory [116] to motive cloud users providing truthful assessments for cloud services.

   (b) In cloud environments, service performance may vary substantially and frequently due to the dynamic nature of cloud services. Thus, continual assessments over time are needed to effectively reflect the dynamic performance of services. In order to motivate a cloud user providing continual assessments, an effective incentive mechanism should be designed, in which the cloud user can be paid if it provides assessments on schedule. However, such a simple mechanism cannot prevent a user from "free-riding" (i.e., providing arbitrary assessments) [97, 193]. Moreover, sometimes an honest user could also provide arbitrary assessments in order to obtain monetary rewards when it does not really know the real performance of cloud services. Such arbitrary assessments may be erroneous and misleading, and therefore greatly affect the effectiveness of service evaluations. To avoid the submission of such arbitrary assessments, an effective incentive mechanism should motivate users to always tell the truth, i.e., allowing honest users to provide uncertain assessments to express their uncertainty about service performance when necessary. To solve this problem, our proposed incentive mechanism is compatible with uncertain assessments, and thus effective for eliciting continual and truthful assessments.

1.4 Roadmap of the Thesis

The thesis is structured as follows:

Chapter 2 introduces the background knowledge of cloud computing, and proposes a generic procedure of service selection in cloud environments. After that, a systematic literature review of cloud service selection is presented in terms of the proposed generic procedure.

Chapter 3 presents the cloud service selection framework and model based on both subjective assessments and objective assessments. This chapter includes our papers published at IEEE SCC 2013 [131] (refer to the publication list on Pages ix and x).

Chapter 4 presents the improved model of cloud service selection, which takes contextual assessments into account. This chapter includes our papers published at IEEE ICWS 2014 [134] and AAMAS 2014 [133].

Chapter 5 presents the approach of evaluating cloud users' credibility of providing assessments. This chapter includes our papers published at IEEE ICSOC 2014 [136].

Chapter 6 introduces *CCCloud*, a combined model of cloud service selection considering both contextual assessments and users' credibility. This chapter includes our papers published at TSC in 2015 [135].

Chapter 7 introduces the uncertain assessment compatible incentive mechanism for eliciting continual and truthful assessments of cloud services. This chapter includes our paper published at AAMAS 2016 [132] and our paper submitted at IEEE ICWS 2016.

Finally, Chapter 8 concludes this thesis.

Chapter 2

Literature Review

In recent years, cloud computing has become the most popular paradigm for storage or service solutions. Due to the diversity of cloud services, it is usually hard for consumers to select the most suitable service. Thus, many efforts have been made in the literature for effective cloud service selection.

In this chapter, the background knowledge of cloud computing is first introduced. Then, a generic procedure of cloud service selection is proposed based on our survey of cloud service selection. The aim of proposing such a procedure is to classify and summarize all the related work in the research of cloud service selection, and identify open issues in this area. After that, a systematic literature review of cloud service selection is presented based on the proposed procedure. In addition, as introduced in Chapter 1, an incentive mechanism for eliciting continual and truthful assessments of cloud services is designed in our proposed approaches for actively improving the accuracy of cloud service selection. Thus, the related techniques of incentive mechanism design is introduced in this chapter.

This chapter is organized as follows:

- Section 2.1 introduces the main techniques applied in cloud computing, and the service models and the deployment models of cloud services.

- Section 2.2 proposes a genetic procedure of cloud service selection, which consists of five steps: 1) defining the purpose of cloud service selection, 2) identifying the roles of service selection participants, 3) service criterion modeling, 4) service criterion evaluation, and 5) service selection execution.

- Section 2.3 presents a survey of cloud service selection. The contributions of the related studies are classified and summarized according to the proposed generic procedure.
- Section 2.4 introduces the related techniques and work of incentive mechanism design.
- Section 2.6 summaries the contents in this chapter.

2.1 Background Knowledge of Cloud Computing

The essential characteristics of cloud computing are *on-demand self-service, broad network access, resource pooling, rapid elasticity* and *measured service* [109]. In order to achieve the above characteristics, two core technologies are applied in cloud computing. The first one is *Service-oriented Architecture* (SOA) [35]. In SOA, a task is accomplished by integrating a range of sub-tasks with independent functions. A sub-task can be considered as a “Service” rendered to consumers. A service can also be composed of other services, but is still a “black box” from the view of consumers [122]. A service in SOA is allowed to be composed and executed repeatedly. Hence, a wide diversity of flexible services can be created and provided to consumers in the cloud environments based on SOA. The other core technology for cloud computing is *“Virtualization”*. Through the logical abstraction, physical computing resources are represented to virtual devices, through which users can easily carry out their computing tasks using the resources without worrying about hardware deployment. A user can access a remote virtual machine like a local computer through the Internet. And all kinds of hardware (e.g., CPU, memory, storage, network, etc.) can be virtualized in cloud environments.

From the perspective of cloud service providers, a cloud environment is composed of three layers: Physical Resource Layer, Resource Abstraction and Control Layer and Service Layer [15]. Physical Resource Layer is the fundamental layer of cloud

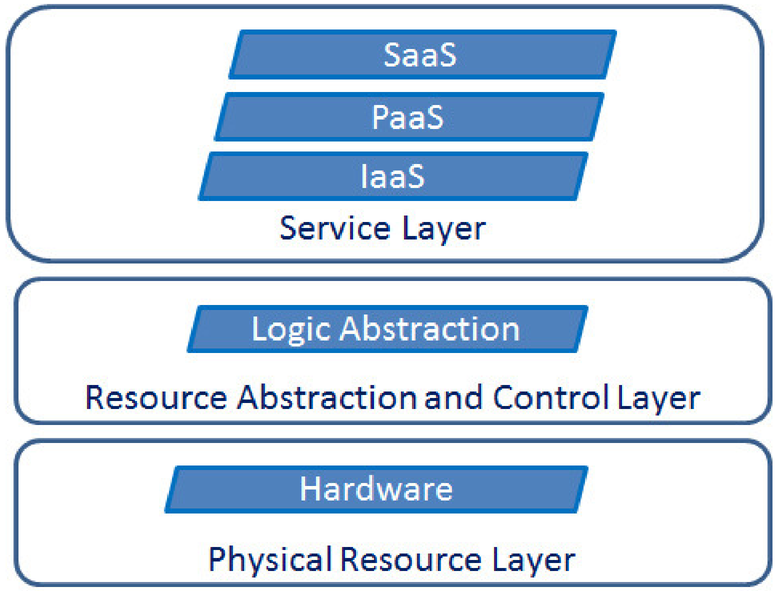


Figure 2.1: Cloud Computing Structure

computing, which contains all the hardware computing resources and facilities of establishing cloud environments. Above the Physical Resource Layer, the Resource Abstraction and Control Layer mainly provides hardware virtualization and management through software abstraction. Virtual machines are typically established at this layer, through which cloud providers can easily manage physical computing resources, and make cloud users conveniently access the hardware resources. The top layer is the application layer, i.e., the Service Layer, where cloud providers define and provision various services to consumers through user interfaces. Fig. 2.1 illustrates the structure of cloud computing.

2.1.1 Cloud Service Models

Based on the thinking from the concepts of SOA and Virtualization, everything in the cloud environments is considered as a service (usually abbreviated as XaaS), e.g., Hardware-as-a-Service, Storage-as-a-Service, Database-as-a-Service, and Security-as-a-Service, Trust-as-a-Service [166, 118]. In general, all cloud services can be typically classified into three service models: Infrastructure-as-a-Service, Platform-as-a-Service and Software-as-a-Service [109] (abbreviated as IaaS, PaaS and SaaS respectively).

- **IaaS** clouds provide all the hardware-related services, e.g., virtual machines, storage, network, firewalls and other fundamental computing resources. Compared to PaaS users and SaaS users, IaaS users have the greatest control of cloud infrastructure. They can develop and deploy nearly any kinds of software, even their own operating systems. The services provisioned from IaaS clouds are usually consumed by PaaS clouds and SaaS clouds. Because IaaS clouds need the deployment of physical computing resource, only a few cloud providers can have the capability to provide IaaS cloud services, e.g., Amazon EC2 [1] and Rackspace [6].

- **PaaS** clouds provide the platforms for users to develop their own applications through pre-defined programming tools, environments and services supported by the PaaS providers. The PaaS users do not need to directly manage the fundamental cloud infrastructure (e.g., operating systems, storage and network), but can possibly adjust the configuration settings of the developing platforms as needed. A PaaS user can use a remote PaaS cloud as a fully configured local computer without managing any hardware settings. Google App Engine [3] and Microsoft Azure [5] are the most popular PaaS clouds.

- **SaaS** clouds provide consumers with the applications, which are developed based on cloud infrastructure and platforms. A SaaS user can access these applications through various user interfaces. The most common and convenient way of building SaaS clouds is to develop web-based applications in cloud environments, so that a SaaS user does not need to install many software products on its own computer, and only need a web browser to access the applications instead, e.g., Google Docs [4]. Compared to the traditional way of applying software, i.e., purchasing a fully licensed software suite, the cost of consuming a SaaS application is quite low since a SaaS application can be shared by many

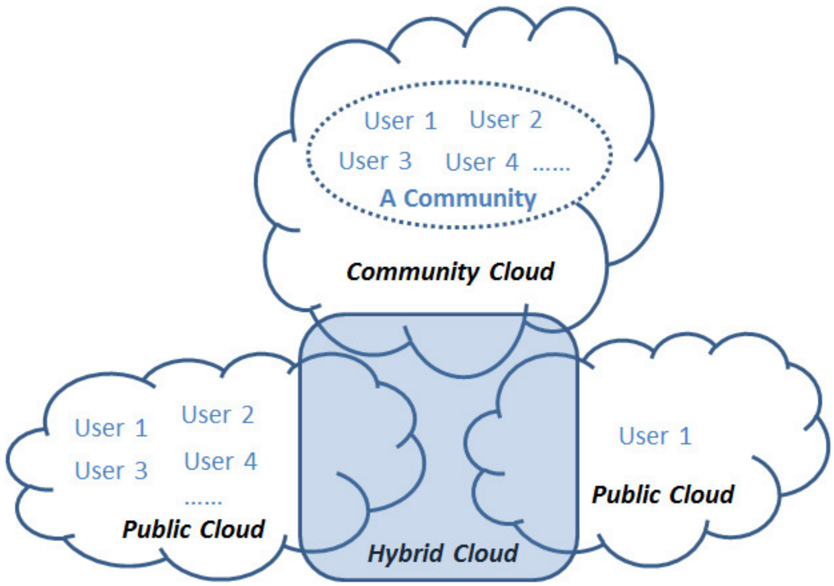


Figure 2.2: Cloud Deployment Models

different consumers at the same time due to the rapid elasticity [109] feature of cloud computing.

2.1.2 Cloud Deployment Models

According to who can access the cloud services, cloud computing has four deployment models: Public Cloud, Private Cloud, Hybrid Cloud and Community Cloud [109]. Fig. 2.2 illustrates the relation among these models.

- **Public Clouds:** a public cloud is the most common way of deploying cloud computing. As the name implies, the users of a public cloud can be anyone, including individuals and organizations, who have purchased the computing resources as needed from the providers. Since the resource pools of a public cloud are shared by many public users, a public cloud can be easily accessed with low costs according to the pay-as-you-go model [11]. Due to this advantage, public clouds are usually chosen by individuals or small businesses. However, just because of such public resource sharing, the public cloud users face various potential security or privacy risks. Hence, public clouds are usually consumed by the users wanting low costs without the needs of high-level security or privacy requirements.

- **Private Clouds:** compared to the public resource sharing of public clouds, a private cloud is only deployed for a particular organization, thus it is more secure than a public cloud because of the exclusive use, but more costly since self-management or maintenance may be needed for a private cloud user. According to the different requirements of consumers, a private cloud can be managed by the organization itself or a third party. The third-party hosted private clouds are usually cheaper than the self-managed private clouds since the expertise of cloud management is required for the latter ones.

- **Community Clouds:** the concept of community clouds is between both of the concepts of public clouds and private clouds. A community cloud can be accessed and shared among a group of several users (usually organizations) who have the similar concerns, e.g., similar tasks, requirements or functionalities. Such a group of users can be considered in the same community. The concept of community clouds further balances the resource cost savings and security and privacy risks in cloud environments.

- **Hybrid Clouds:** the concept of hybrid clouds refer to the combination of two or more of the other three cloud deployment models. It succeeds the advantages of the other models, and bring greater flexibility for satisfying various requirements of consumers. For example, the core tasks of a cloud user's business, requiring high-level security and privacy, can be executed in a private cloud operated by itself for minimizing the possible risks, and the usual tasks can be executed in a public cloud environment for minimizing costs.

It is due to the diversity of the service models and the deployment models of cloud computing that cloud service selection is more challenging than traditional web service selection, and there is no service selection solution which can be applicable in every type of cloud services. Hence, in order to cover all the situations of service selection in cloud environments, we propose a generic procedure of cloud service selection, and survey and classify the existing works according to the proposed generic procedure and the applied technologies.

2.2 A Generic Procedure of Cloud Service Selection

In order to cover all the existing approaches or models in the literature, we propose a generic procedure of cloud service selection, which includes five steps: 1) defining the purpose of cloud service selection; 2) identifying the roles of service selection participants; 3) service criterion modeling; 4) service criterion evaluation; and 5) service selection execution. We expect to map all the existing works of cloud service selection into the generic procedure, and hence identify the research focuses of cloud service selection and the related open issues.

2.2.1 Defining the Purpose of Cloud Service Selection

In the literature, the purposes of cloud service selection can be generally classified into two kinds: selecting *the best service* and selecting *an optimal composition of services*.

- **Selecting the best service:** the outcome of this kind of cloud service selection is typically the best alternative service or a service ranking according to the requirement of a cloud consumer asking for service selection. The selection process is usually based on the performance of the multi-faceted aspects of alternative services, and is thus modeled as the multi-criteria decision making (MCDM) problem. The consumer can determine the importance level for every performance aspect, and thus the selection result can reflect the consumer' customized requirement. The common approaches for the MCDM problem include AHP, ANP, weighted sum and fuzzy decision making, etc [39].

In addition, the best service selection can also be carried out based on the trustworthiness evaluation of services or service providers. In the literature of cloud service selection, the trustworthiness refers to how satisfactory or reliable the provisioned quality of services is from consumers' perspectives. In these cases, trust models or trust-based approaches are applied, e.g., [54, 47, 105, 167].

The results of the best service selection are usually computed for end-consumers. In some cases, the selection results can also be taken as input for service composition executed by application cloud providers.

- **Selecting an optimal composition of services:** in this kind of cloud service selection, a consumer usually needs to select a group of composed services in order to meet his/her requirements on complex functionalities. One simple example is that, a travel agency provides a trip plan service, which is composed by other sub-services, including booking a flight, booking a hotel, renting a car, ordering admission tickets, etc. Then an optimal trip achieving some objectives, e.g., minimizing the travel duration, is planned for a consumer according to his/her requirements. Such requirements can be considered as the constraints in the service composition, e.g., the total cost must be lower than 1000 dollars. Hence, the service composition problem can be modeled as the knapsack problem [147], which is typically solved through optimization-based approaches, e.g., [77, 128, 20].

  The optimization process is mostly based on the performance quality of services. In addition, the trustworthiness of services or service providers is taken into account in some studies, e.g., [56, 182]. The studies of service composition in general service-oriented environments usually focus on the composition structures, e.g., sequence, parallel and loop. However, in the literature, the studies of cloud service composition focus on either structural or non-structural composition. For example, an application cloud provider wants to optimize the computing resources, which may be provisioned by different infrastructure cloud providers, for satisfying various consumers' requirements. In this case, the invoking relations among service components do not affect the whole service performance since every service component serves different groups of consumers separately.

2.2.2 Identifying the Roles of Service Selection Participants

In general, service providers and service requesters are the two basic roles in all service selection problems. In cloud environments, service selection not only happens between cloud providers and cloud consumers but also between data centers and cloud providers. Under the latter context, data centers are the service providers; cloud providers become the service requesters. The cloud providers need to select suitable data centers to support satisfactory cloud deployment with minimum costs. Furthermore, due to the layers of cloud service models, a cloud provider can be the service requester for some lower-layer cloud providers (e.g., IaaS or PaaS providers), and serve the higher-layer cloud consumers (e.g., SaaS consumers). Hence, the participants of cloud service selection can be classified into four groups: end consumers, application cloud providers, infrastructure cloud providers and data centers.

- **End consumers** are the ordinary consumers using cloud-based applications. They first select the cloud services having the required functionalities, and then concern their non-functionality performance, such as service availability and response time. Thus, the service selection for end consumers is usually based on the MCDM approaches considering both functional and non-functional attributes of cloud services.

- **Application cloud providers** offer application services to end consumers and require cloud deployment services from infrastructure cloud providers. They also require functional services from other application cloud providers, and compose them for complex functionalities. Thus, the service selection for application cloud providers may consider both multiple criteria evaluation and optimal service composition.

- **Infrastructure cloud providers** are in charge of cloud deployment, and hence needs to select suitable cloud hardware and facilities. The cloud hardware are typically provisioned by worldwide data centers. In order to satisfy the different levels of service quality for different kinds of consumers, infrastructure cloud providers need to consider the contexts of the data centers, e.g., geographical location and peak hour loads. In the literature, some studies have been proposed from the perspective of location-aware selection of data centers, e.g., [58, 128, 129]. Due to the physical features of hardware, context-aware service selection should be paid more attention for infrastructure cloud providers.

- **Data centers** provide the fundamental hardware and resources for cloud deployment. Every data center has its resource capacity and serves consumers from a particular region because of the consideration of the latency of resource delivery. In order to maintain the quality of services for consumers around the world, an infrastructure cloud provider needs to dynamically allocate and distribute computing resources from different data centers. Hence, a consumer can experience a stable cloud service regardless of his/her location.

Fig. 2.3 illustrates the relations between these roles in cloud service selection. As introduced above, cloud service selection may be quite different for different kinds of cloud providers and consumers, and thus needs to be studied under particular contexts. In the literature, most of the studies of cloud service selection only consider the general provider-requester scenario, and do not specify the application contexts of their studies.

2.2.3 Service Criterion Modeling

An essential issue of cloud service selection is how to determine the necessary service criteria, which decide whether an offered cloud service can satisfy both business and

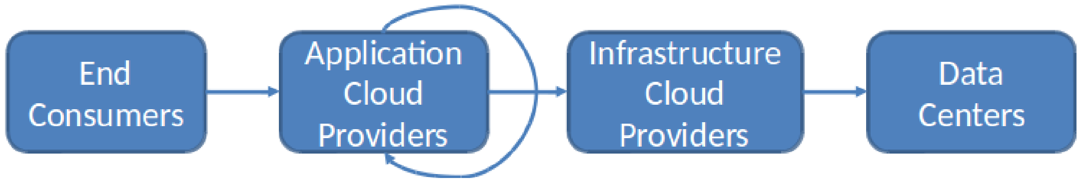


Figure 2.3: The Roles in Cloud Service Selection

technical requirements of cloud consumers. The requirements includes both functional and non-functional aspects of cloud services. To be brief, a functional requirement refers to what specific behavior a service can do; a non-functional requirement refers to how well a service can behave.

Due to the diversity and complexity of cloud services, determining cloud criteria for global use is very challenging. Even so, many efforts have been made in the literature. One widely accepted criterion model for cloud services is the cloud service measurement index (SMI) proposed in [146], where cloud criteria are grouped into seven major categories:

- **Accountability** refers to how much a cloud consumer can rely on a cloud provider, or put another way, the trustworthiness of a cloud provider to provide satisfactory services.
- **Agility** represents the ability of a cloud service to quickly respond to consumers' requirements of changing or extending the use of the service.
- **Assurance** indicates the ability of a cloud service to provide the expected functions as specified.
- **Financial** indicates how much money should be paid for accessing an agreed cloud service.
- **Performance** represents how well a cloud service can perform as expected.
- **Security and Privacy** guarantee the secure management of consumers' data in a cloud, with unauthorized parties being unable to access these data.

Category	Attribute
Accountability	Provider contract / SLA verification
	Compliance
	Ease of doing business
	provider certifications
Agility	Scalability
	Portability
	Elasticity
Assurance	Availability
	Reliability
	Resiliency / fault tolerance
Financial	On-going cost
	Acquisition and transition costs
Performance	Service response time
	Functionality
	Interoperability
Security and Privacy	Access control
	Data privacy and data loss
	Data integrity
Usability	Accessibility
	Learnability
	Suitability

Table 2.1: Prioritized SMI Attributes

- **Usability** represents how easily a cloud service can be used. A cloud provider should provide sufficient business and technical support to various consumers.

In each major category, there are several attributes to represent the specific performance aspects. In [146], a tree structure is proposed to present 51 specific attributes in total. Table 2.1 [146] shows some primary attributes in every major category. Note that, most of the attributes shown in Table 2.1 are abstract, and the detailed modeling processes of these attributes are not presented in [146]. Another widely referenced SMI is *SMICloud* [45], which focuses on numerically expressing the performance attributes, including *service response time, sustainability, suitability, accuracy, transparency, interoperability, reliability, stability, cost, adaptability, elasticity* and *usability*. For each attribute, an intuitive calculation is carried out, the results of which can be taken as input for any cloud selection approaches.

It should be noted that these proposed SMIs can only represent the general performance of cloud services, and some of the performance attributes can hardly be expressed via quantitative forms, e.g., security and privacy. In practice, how to accurately model these attributes is still challenging. A cloud consumer may have more specific requirements for a cloud service. The result is that the proposed SMIs are not so likely to be applied in practice since they can hardly reflect the concrete needs of consumers. Hence, before cloud service selection, the core performance attributes should be first selected and modeled according to consumers' personal requirements.

In the literature, many efforts have been made to model and evaluate the concrete cloud performance attributes for service selection. Tang and Liu [153] propose a framework for security control of SaaS clouds, which mainly contains five security aspects: *Function, Auditability, Governability* and *Interoperability*.

Furthermore, a consumer' preference on each performance attribute can cause quite a different result of cloud service selection. A consumer should be allowed to set the importance weight or QoS constraint for each attribute before carrying out service selection. However, accurately setting such weights or constraints is still tricky for ordinary consumers. That is because quantifying importance weights or required constraints usually needs expert judgment. However, ordinary consumers may be lack of such expertise, and have subjective vagueness or bias in some cases. Hence, service criterion modeling should support consumers to accurately identify their preference.

2.2.4 Service Criterion Evaluation

After service criterion modeling, how well every concerned attribute performs should be further determined in the step of service criterion evaluation, i.e., determining QoS value of every attribute. In the literature, cloud service criterion evaluation can be generally classified into three types:

- **QoS description:** in early studies of cloud service selection [140, 52, 115], QoS values are assumed to be extracted from service description which is commonly included in SLA. However, cloud providers in practice may always have incentives to provide unsatisfactory QoS in order to reduce cost. And detecting SLA violation is costly for ordinary consumers. Thus, recent studies trend to consider extracting QoS values from other ways rather than service description.

- **Service-side QoS evaluation:** in the real-world situations, the most common way of getting QoS values is to apply the key performance indicators (KPI), which are usually provided by service providers, and executed at service side, e.g., Amazon CloudWatch[1]. However, service-side QoS evaluation cannot effectively reflect end-users' experience of service quality. Furthermore, the KPIs provided by cloud providers are usually insufficient to present all the concerned aspects of service performance for consumers, and may not be fully trusted since providers may cheat consumers by presenting excellent but fake performance records.

- **End-user-side QoS evaluation:** in order to solve the problem in service-side QoS evaluation, many studies focus on QoS evaluation from end-user side, e.g., [54, 118, 167, 32, 101]. Such an evaluation is based on the collection of user feedback or testing reports. The QoS assessments from end-users may be either subjective or objective. Subjective assessments refer to the assessments extracted from users' subjective judgements about service quality, e.g., ratings. Objective assessments refer to the assessments acquired from quantitative performance monitoring or testing at the end-user side. These subjective or objective assessments are taken as input for cloud service selection approaches.

  Note that, in end-user-side QoS evaluation, assessment contexts should be taken into account, i.e., under what situation an assessment is made. That is because cloud consumers under different contexts may have quite different experiences of cloud performance. For example, consumers in different geographic locations may have different judgments due to the latency of service delivery. In another

[1] aws.amazon.com/cloudwatch

example, cloud performance may be different during peak hours and non-peak hours due to the work-load variations, and thus, cloud assessments collected from end-users during various moments may be quite different. Considering this issue, context-aware QoS evaluation should be further studied when carrying out end-user-side service evaluation. But only few studies [134, 135] in the literature focus on it.

Based on the above types of cloud service criterion evaluation, QoS assessments can be classified into two groups:

- **Objective assessments** are used for the performance attributes which can be represented through quantitative values, e.g., 95% for service availability and $4ms$ for response time. Due to the complexity of cloud services, some performance attributes cannot be observed directly, e.g., cloud reliability, scalability and elasticity. Furthermore, in some cases, some specific performance aspects needs to be evaluated according to consumers' particular requirement, e.g., CPU or storage speed. Thus, many benchmark approaches or tools are specially designed and proposed for cloud QoS evaluation [23, 14, 27, 84, 40, 24, 65, 48, 65].

- **Subjective assessments** are usually more convenient to be acquired and collected than objective assessments due to the absence of the cost of quantitative performance evaluation, e.g., extra software development and installation. Subjective assessments are used for the performance attributes which cannot be easily quantified but can be customarily judged through consumers' subjective experiences in the daily uses of services. For example, a consumer can judge whether the functions of a service are easy to handle, or whether a service provider provides sufficient after-sale supports for its customers. In addition, for some performance attributes, if a way of quantitative evaluation cannot be widely accepted for various consumers, consumers' subjective experiences could be used as the assessments of these attributes for different situations, e.g., service security and privacy.

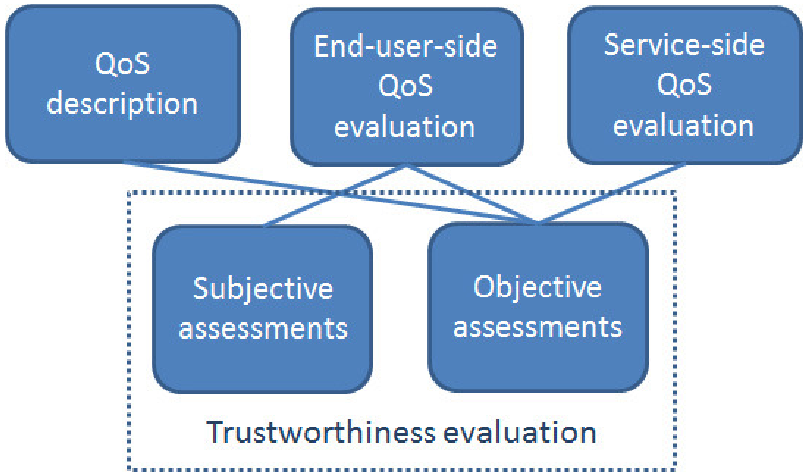


Figure 2.4: Service Criterion Evaluation

Note that, a big issue in service criterion evaluation is the trustworthiness of performance assessments. As introduced above, the objective QoS definitions from QoS description may be violated; objective assessments from service-side evaluation may not sufficiently reflect end-users' experiences; and subjective assessments may contain much noise, which includes consumers' subjective biased assessments or even malicious assessments on purpose. Hence, some research [99, 101, 136, 135, 163, 117, 118] in regard to the trustworthiness evaluation of performance assessments has been undertaken. However, most of the studies of cloud service selection tend to assume that the acquired QoS assessments can be fully trusted.

Fig. 2.4 illustrates the relations between service criterion evaluation types and QoS assessment types. Objective assessments of cloud services can be acquired in all the three evaluation types; and subjective assessments are usually extracted from end-user-side QoS evaluation. And the trustworthiness of both subjective assessments and objective assessments needs to be evaluated according to the particular requirements of consumers who request cloud service selection.

2.2.5 Service Selection Execution

Based on the above steps, a consumer confirms his/her purpose and role of service selection, and determines what performance attributes need to be necessarily considered and his/her preference of these attributes. After obtaining trustworthy performance assessments, the final service selection is ready to be executed. In the literature, cloud service selection can be carried out through four types of approaches: *MCDM-based selection*, *optimization-based selection*, *trust-based selection* and *description-based selection*. The details of these approaches are introduced in the next section.

Fig. 2.5 presents the whole generic procedure of cloud service selection. We expect to include all the existing works of cloud service selection based on this generic procedure, so that the research contributions can be summarized, and the further issues can be identified.

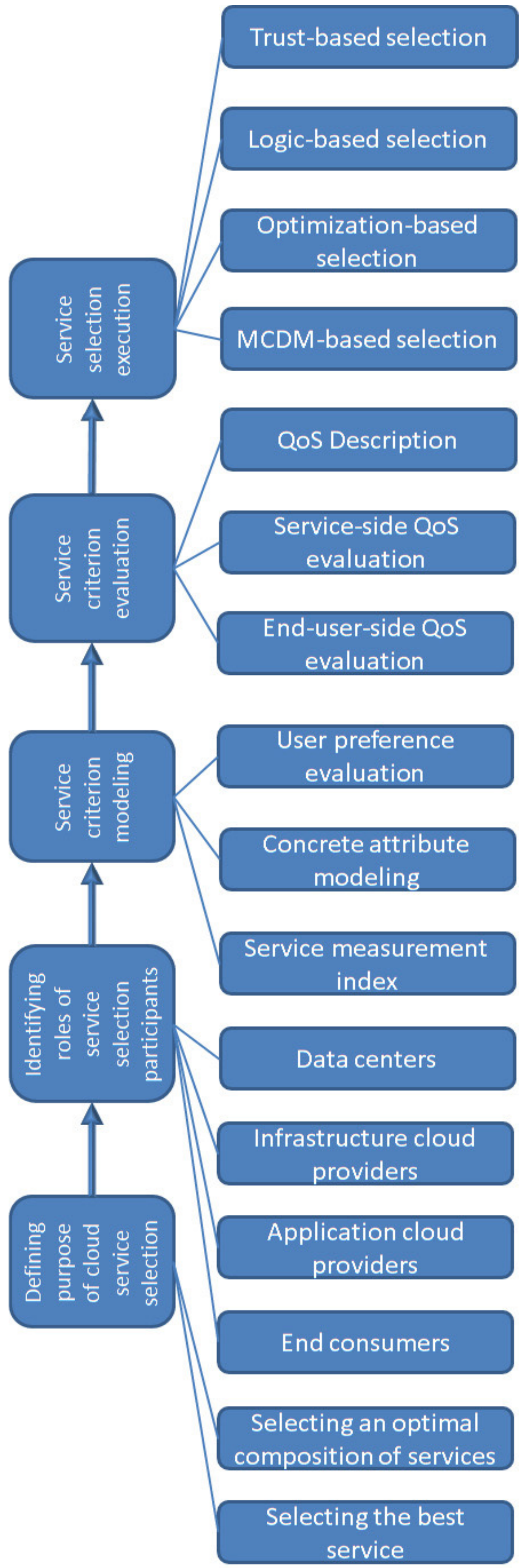


Figure 2.5: The Generic Procedure of Cloud Service Selection

2.3 Techniques of Cloud Service Selection

In this section, we introduce the techniques of cloud service selection, and then categorize the existing approaches based on our proposed generic procedure. Through such a categorization, the outstanding contributions of the existing studies are summarized.

2.3.1 MCDM-based Selection

Multiple-criteria decision making [157] is the most common technique applied in cloud service selection. In this scenario, the overall performance of a cloud service is expressed via the aggregation of a set of finite attributes, each of which represents one performance aspect. Hence, the main issues of MCDM-based selection are the comparison of the performance quality of every attributes among all the alternative services, and the aggregation of the performance attributes according to consumers' requirements.

2.3.1.1 QoS Prediction

As introduced in the proposed generic procedure, the QoS assessments applied in cloud service selection are typically obtained through three ways: pre-defined service descriptions, ordinary consumers' judgments or quantitative monitoring and testing. However, in some cases, such assessments may be missing, e.g., new cloud services appear or cloud performance records are incomplete. To solve this problem, some QoS prediction approaches are proposed in the literature of cloud service selection. The most popular techniques applied in QoS prediction include collaborative filtering (CF), clustering and matrix factorization.

In [78], Karim *et al.* propose a performance prediction approach of cloud services from end-users' perspective based on collaborative filtering [67]. In this work, it is assumed that a typical cloud solution is composed of two layers of services: 1) software services satisfying consumers' functional requirements, and 2) infrastructure services supporting the software services on service performance. Thus, the quality of services an end-user can experiences depends on the combination of the services from the two layers. If a particular combination of services has not been achieved, the prediction approach can be applied to estimate the end-performance of such a combination based on the QoS of similar services which have been combined before.

In [195], Zheng *et al.* propose a framework of ranking cloud services according to a particular performance attribute, e.g., response time. The proposed work focuses on computing the personalized ranking for particular users according to their preferences. Based on users' historical ratings for services, the ranking similarity between users is computed for determining similar users. Finally, a user's preference for an alternative service can be predicted through a Greedy Order Algorithm according to the experiences of his/her similar users. In [194], Zheng *et al.* extend their approach. In the extended version, the ranking prediction can be achieved without predicting the exact values of missing QoS, through which the ranking accuracy can be improved due to the less effect on users' subjective bias.

In [154], Tang *et al.* propose a QoS prediction approach for cloud services. The proposed work focuses on solving the data sparsity problem for cloud QoS prediction. Cloud users are first clustered based on their location. Then the data smoothing technique is applied to fill missing QoS values using the average values of the items from the same user cluster. The final QoS prediction is carried out by combining user-based and service-based collaborative filtering.

In [184], Yu propose a personalized recommendation approach for cloud services using matrix-factorization-based clustering. The users who have the similar historical experiences are assumed to have the similar cloud-related features in common. Likewise, the cloud services providing similar QoS to users are also considered to have some common features. Based on these features, cloud users and services are clustered into a group of communities, which are used to predict users' future QoS experiences.

2.3.1.2 MCDM-based Approaches

After obtaining sufficient assessments of alternative cloud services, the decision-making process would be carried out. In the literature, various MCDM techniques are applied for cloud service selection, the most popular techniques of which include Analytic Hierarchy Process/Analytic Network Process (AHP/ANP), weighted sum, Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), logic inference, fuzzy decision making, etc. [39].

In [161, 137, 160, 159], Rehman *et al.* propose a framework supporting IaaS cloud selection. In this framework, it is assumed that the performance of a IaaS cloud is monitored and recorded through a service-side status checker installed on the virtual machine accessed by IaaS users. Then the cloud performance reports are gathered and managed by a centralized QoS repository for further service evaluation and selection. Furthermore, they propose a time-aware selection approach based on different MCDM techniques, e.g., TOPSIS and AHP. It is argued that the most recent performance reports can more accurately reflect the current performance of cloud services, and thus are more helpful for cloud consumers. Based on this consideration, the QoS records of alternative clouds are logistically decayed according to the time distance from the times pot, at which the selection decision is required to be made. A case study is presented to demonstrate the effectiveness of the proposed approach based on the real data of CPU, memory and I/O collected from Amazon EC2. They conclude that different MCDM techniques may not lead to the same decision results in the same case.

In their further study [162], they propose a service selection approach considering the dynamic nature of cloud computing. In this work, a cloud service is evaluated through time-aware QoS monitoring. The continual QoS assessments of alternative services are divided into many time windows according to their submission time. The service ranking is carried out in parallel with every time window. Then the ranking values in each time window are weighted to compute the aggregated ranking of all alternative services. The weights of the sub-rankings are decayed with time in order to guarantee the freshness of service evaluation. At last, the real dataset collected from *CloudClimate*[2] is applied to validate the proposed approach.

In [49], Godse and Mulik propose a AHP-based approach of SaaS cloud selection. The SaaS product selection is modeled based on a hierarchy consisting of five main factors (Level-1), including *Functionality, Architecture, Usability, Vendor Reputation* and *Cost*, and sixteen attributes (Level-2). The proposed hierarchy is applied in a case study of *Sales Force Automation* products.

In [71], Gonçalves *et al.* propose a AHP-based cloud deployment selection driven by the non-functional performance attributes, including *efficiency, cost* and *scalability*. The proposed approach is employed in a real-world case (WordPress deployed in the Amazon cloud).

In [7], Achar and Thilagam propose a cloud service selection approach based on the Service Measurement Index [146]. The selection procedure consists of three steps: 1) identifying concerned criteria according to the SMI; 2) determining the criterion weights; and 3) ranking alternative services using TOPSIS. A simulation experiment is carried out to evaluate the proposed approach through the CloudSim toolkit[18].

In [113], Mu *et al.* focus on the study of precisely estimating users' preferences for the performance attributes applied in the decision making of cloud service selection. The users' preference represented by the importance weight set for every attribute is assumed to be classified into two types: subjective weights and objective weights. The subjective weights are directly set by users. In order to deal with the vagueness contained in the subjective weights, fuzzy weights are applied to express users' uncertainty. On the other hand, objective weights are computed from the preference history of the same service request through Rough Set [125] if users did not set the subjective weights. Then the subjective weights and objective weights are combined to reflect users' preferences. The service selection process is carried out using TOPSIS.

In [142], Sahri *et al.* focus on the particular selection of cloud-based databases. The selection is supported by an ontology-based framework named "DBaaS-Expert",

[2] www.cloudclimate.com

and carried out through AHP based on *Quality of Service, Capacity of Service* and *Cost of Service*.

In [149], Sun *et al.* study the issue of uncertainty in cloud service evaluation based on users' subjective judgements, which are expressed through triangle fuzzy numbers. The operations of addition, subtraction, reciprocal and multiplication are defined based on the fuzzy set. Through an ontology, the weights of selection criterion are processed through a fuzzy-based AHP algorithm, and then the alternative services are ranked through a fuzzy-based TOPSIS algorithm. In their later study [150], two types of inter-relationships between criteria are taken into account, i.e., supportive and conflicting. The relations of criteria are modeled through an interactive and interpretive structural network, based on which the final selection is carried out via ANP [141]. The similar works of cloud service selection based on fuzzy sets using AHP are proposed in [152, 59], and another ANP-based work which considers the inter-relationships among performance attributes is proposed in [25].

In [168], Wang *et al.* propose a logic-based model for cloud service selection. The proposed model consists of three steps: 1) the performance of the alternative services is first evaluated through fuzzy synthetic decision according to cloud consumers' personalized preferences; 2) then the quantitative performance values of the services are converted into uncertainty levels through the Cloud Model [86]; and 3) the outputs from Steps 1&2 are taken as input into a Fuzzy Logic Control scheme. Through rule-based fuzzy inference, the alternative services are ranked for selection.

In [145], Shivakumar *et al.* propose a fuzzy-based cloud service selection approach, in which different performance attributes are expressed by different fuzzy membership functions, and the final decision is extracted from the aggregation of all the attributes through a pre-defined fuzzy "and" operator.

In [46, 45], Garg *et al.* propose a framework of ranking cloud services, in which a service measurement index (SMI) is presented to quantitatively measure the performance quality of services. They propose a range of formulas to model the general performance attributes of cloud services, which may not be easily expressed in quantitative forms, e.g., sustainability, suitability, reliability, stability, etc. Based on this quantitative modeling, the alternative services are ranked through AHP.

In [143], Saripalli *et al.* introduce a methodology of cloud service selection based on MCDM and a cloud taxonomy. Instead of applying technical selection approaches, the Wide-band Delphi Method [121], which is a structured communication technique based on iteratively asking opinions from experts, is recommended in this work.

In [79], Karim *et al.* focus on cloud service selection from end users' perspective. In their work, a cloud service is assumed to be composed of two layers of services, i.e., the SaaS layer and the IaaS layer. cloud consumers' requirements of service QoS specifications are first mapped into the SaaS layer, and then mapped into the IaaS layer through a hierarchy-based model. The services in the IaaS layer are ranked through AHP. Through such mapping and ranking processes, the authors argue that an application cloud provider can select the suitable infrastructure cloud services which can exactly satisfy end-users' QoS requirements. A similar mapping mechanism is proposed based on QoS ontology in [80], which is implemented via OWL-S (Web Ontology Language for Services) [53] and SWRL (Semantic Web Rule Language) [61].

In [110, 111], Menzel *et al.* propose a framework for cloud infrastructure service comparison, and present a generic procedure for immigrating old-school business into clouds. The proposed framework and procedure are designed to be compatible to any specific requirements and preference as well as any MCDM techniques, e.g., AHP and ANP. A prototype is proposed to validate the feasibility of the proposed framework.

Qu *et al.* propose a range of studies in cloud service selection [131, 134, 133, 136, 135]. In these works, they argue that, cloud service evaluation and selection are typically based on either subjective assessments from ordinary consumers (e.g., ratings) or objective assessments through quantitative performance monitoring and testing. Either type of assessments has the corresponding limitation in comprehensively reflecting cloud performance. Thus, they propose a framework of cloud service selection based on the comparison and aggregation of both subjective assessments and objective assessments [131]. In their further study [134], they focus on personalized and credible service selection, and take the contexts of assessments and users' credibility into account. They argue that the cloud assessments generated under a context are more reliable for the cloud requestors under the similar context. In this work, two assessment context features (*location* and *time*) are considered for determining more accurate assessments in service selection. Furthermore, a theoretical model is proposed in [135] to evaluate the credibility of cloud users who offer assessments The proposed model can thus effectively reduce the impact caused by biased assessments and noisy assessments in cloud service selection.

In [148], Srivastava and Sorenson argue that users' preference of QoS performance usually varies non-linearly with the real QoS performance, and thus propose a service selection approach based on users' subjective judgments instead of objective QoS value. Their study focuses on finding out an accurate relation between users' ratings and actual QoS performance, which is solved by the mid-level splitting method [19]. The main drawback of this work is that the accurate relation needs to be learned from substantial queries to users. In addition, users' subjective vagueness and bias are not taken into account.

2.3.2 Optimization-based Selection

In service selection, optimization-based selection refers to finding the most suitable service or service group, which can achieve one or more objectives, e.g., minimizing costs, maximizing performance quality, etc. Moreover, the optimization objectives typically need to be achieved under specific constraints, which are usually required by service users, e.g., consumption budgets. Hence, in this scenario, service selection can be modeled as the *Selective Multiple Choice Knapsack Problem* which has been proved to be a NP-hard problem [166].

In the general service-oriented environments, optimization-based selection is widely studied in the field of service composition. Many efforts have been made in this area, e.g., [16, 55, 63, 9, 64, 95, 98, 158, 186, 185, 10]. In these studies, services are structurally composed for a single task. For each task component, there are multiple functionality-equivalent alternative services. According to the capacities of the alternative services and users' demands, an optimal service composition plan is determined. In addition to structural service composition, in cloud environments, optimization-selection can also be applied in non-structural service combination. For example, an infrastructure cloud providers need to select multiple data centers serving worldwide consumers in order to minimize service response time. In this case, the structural relationships among the data centers are not important for the final performance of the provider.

In [58] and [57], Hans *et al.* propose a heuristic approach for an infrastructure cloud provider selecting a set of data centers with the goal of minimizing the total cost. The provider aims to provide services to consumers distributed in different geographical locations. In order to reduce the latency of service delivery and maintain the quality of services, the provider expects to select multiple data centers in different locations, each of which serves the consumers in the near location. The data center selection problem is modeled as an optimization problem solved using integer programming. In order to improve the computation efficiency, a priority-based heuristic algorithm is proposed based on liner programming relaxation [50].

A similar study of data center selection in cloud environments is presented in [129] and [128]. Compared to Hans *et al.*'s work, Qian *et al.* consider more characteristics in data center selection. They argue that, in addition to data center distribution and cloud user distribution, dependencies among application components should be studied due to the complex composition of distributed applications. Moreover, the locations of related applications should also be considered in order to reduce the communication latency between applications. The whole selection problem is modeled into a graph including data centers, cloud providers, application components and cloud users under specific constraints (e.g., resource capacities, geographical distances, cost, etc.). A heuristic approach is proposed to achieve the global goal of selection, i.e., minimizing both the total implementation cost and service delivery distance.

In [170], Wang *et al.* propose a SaaS cloud recommendation approach in multi-tenant environments, where cloud consumers have different and multi-dimensional QoS requirements, while cloud providers have their own optimization goals, e.g., minimizing costs. The main aim of the proposed approach is to improve the efficiency of service selection. In this work, cloud tenants are first clustered according to the similarity of their requirements on multiple QoS attributes. Then the alternative services are clustered based on the tenant clusters. The alternative services in a service cluster are ranked based on their utility values computed through weighted QoS attributes. At last, a greedy algorithm [29] is applied to find the optimal service composition for every tenant cluster according to the service rankings in every service cluster. The experimental results demonstrate that the cluster-based approach can greatly reduce the computation time, while maintaining the success rates compared to other related approaches.

In [169], Wang *et al.* propose a distributed framework of cloud service selection based on cloud brokers. The framework consists of three layers: user layer, broker layer and service layer. In the user layer, cloud users submits their requirements of cloud services to the suitable brokers via user agents. The cloud brokers in the broker layer are in charge of collecting and updating cloud performance information. According to the types of registered cloud services, cloud brokers are clustered in order to reduce the service retrieval time. In the service layer, the performance of services is monitored and registered to the brokers. An incentive function is proposed to motivate cloud providers submitting effective performance information to the suitable brokers. Through a set of adaptive learning algorithms, the broker selects the optimal service with the maximum performance-to-cost rate according to the information collected from both users and providers.

In [77], Kang *et al.* focus on modeling cloud users' vague preferences in cloud service composition. They argue that the precise user requirements and preferences required in traditional service composition are hardly achieved in practice. Thus, they propose an approach to learn users' preferences according to their historical ratings of composed service performance. If a user does not have sufficient records for learning, the service composition plans from similar users would be recommended. If a new user does not have any historical record, the most favorite service composition plan would be recommended until the user generates enough records. Note that, the proposed approach only considers three structures of service execution plans, i.e., individual services, sequential services and parallel services.

In [22], Chen *et al.* model cloud service selection as a *multi-objective p-median problem with dynamic demands*, i.e., according to k pre-defined optimization goals, selecting p cloud services from m alternative services for serving n cloud users, which is proven NP-hard [83]. The proposed selection model considers four optimization objectives: minimizing QoS delivery between users and services, minimizing network transmission cost, minimizing service cost and maximizing the total number of provision services. In addition, users' requirements for services can change over time. A genetic algorithm is proposed in order to address the dynamic-demand setting.

In [89], Li *et al.* model cloud service selection as a constraint-based optimization problem which considers two types of constraints: functional constraints defined via WSDL (Web Services Description Language) and non-functional constraints expressed via QoS attributes. The proposed model consider three composition structures: sequence, parallel and switch (i.e. if-then-else structure). Based on these structures, a set of QoS aggregation functions are proposed for four QoS attributes, i.e., *Time, Price, Availability* and *Reliability*. The optimization problem is solved through the standard Particle Swarm Optimization algorithm [126].

In [20], Chang *et al.* propose a probability-based approach for selecting cloud storage providers. They argue that a cloud user needs to duplicate his/her data in multiple storage clouds in order to maximizing the data availability. The novelty of this work is to consider service selection under a particular context (i.e., storage services), and thus the selection objectives are determined according to such a context, which includes minimizing data access failure probability, maximizing data survival probability and maximizing the expected number of survival data blocks when some storage providers fail to work properly. Two dynamic programming algorithms are proposed in terms of different optimization objectives.

In [60], He *et al.* propose a QoS-driven cloud selection model for SaaS cloud providers in the multi-tenant environment. The proposed model is motivated by such a scenario: a SaaS cloud provider wants to provide an application services to multiple consumers. These consumers require the same functionality but different QoS performance. Thus the provider tries to satisfy all consumers' QoS requirements, while achieving its own goals (e.g., minimizing cost or maximizing performance quality). In the proposed model, four types of service composition structures are studied, i.e., *sequence, branch, loop* and *parallel*, based on which the corresponding QoS aggregation functions are defined for the QoS attributes, including *cost, response time, availability* and *throughput*. For solving the optimization problem, the authors analyze and compare three methods: integer programming, skyline and a greedy algorithm, where the greedy algorithm outperforms the other methods.

In [34] and [94], Du *et al.* and Liang *et al.* study the temporal constraint problem in cloud service selection and composition. In their studies, the temporal constraints refer to a kind of time constraints, e.g., a service component is required to be invoked no later than a specified period of time after a prior component accomplishes its task. Thus, the violation of temporal constraints should be detected at run time, and then the corresponding composed services should be re-planned if necessary. To this end, a penalty-based genetic algorithm is applied for dynamical service composition. In this algorithm, a checkpoint strategy is used to detect the violation of temporal constraints. If a violation happens, the fitness of the interim solution in a generation of the algorithm would be reduced in order to ensure finding more optimal solutions in the further generations.

In order to improve the efficiency of cloud service selection, Sudareswaran *et al.* propose a selection approach based on cloud brokers [151]. In their approach, all the related information of cloud services is encoded and indexed. The cloud services with similar characteristics would be clustered for quick service searching. According to consumers' requirements, the services having the closet guarantees to the requesters' queries are selected as the best services through a greedy algorithm according to their Hamming distances.

In [104], Martens and Teuteberg propose a cloud service selection approach based on analyzing cost and risk. They argue that most of the decision factors in cloud selection are related to costs, and thus can be represented by cost values, e.g., adoption costs, maintenance costs, etc. In addition to costs, the risks of consuming cloud services are classified into three aspects: integrity, confidentiality and availability. The final optimization goal is to minimize both the total costs and risks. Several formal models are proposed for cost and risk evaluations. It is argued that various optimization techniques can be applied on the proposed models, e.g., linear or non-linear programming, and genetic algorithms.

Different from most of the studies of service composition, Yang *et al.* [181] focus on cloud service composition at runtime, rather than the service composing-then-executing way. They argue that, during service execution, many unpredictable factors would affect the whole performance of the pre-planned service composition, e.g., alterable service requirements and unstable service performance. Hence, service selection and composition need to be dynamically processed during execution. They model such a problem via Markov decision process [127]. When a service component in a task is unavailable or unstable, the rest of required services would be recomposed based on their utilities. The performance of the following services would be estimated through MDP. Through experimental simulation, the proposed approach can achieve a close-to-optimal solution, while greatly reducing the time cost. Note that, only the sequential composition structure is considered in the proposed approach.

2.3.3 Trust-based Selection

The concept of "trust" has been introduced in many different contexts. In service-oriented environments, Jøsang *et al.* define trust as "the subjective probability by which an individual expects that another performs a given action on which its welfare depends [70]." We define that trust-based selection in cloud environments refers to service selection based on the evaluation of the trustworthiness of cloud providers or services. In trust-based selection approaches, trust models are typically proposed based on either subjective assessments or objective assessments, or the combination of them. Subjective assessments are usually extracted from ordinary consumers' judgements or expert opinions; and objective assessments are usually obtained from quantitative performance evaluation or QoS descriptions.

In [54], Habib *et al.* propose a trust management system in cloud environments, the aim of which is to help cloud consumers find trustworthy cloud providers from vague service descriptions. The system can take both subjective assessments from user feedback and objective assessments from service auditing as input. For each performance attribute of a cloud service, the corresponding trust is evaluated based on the average rating of the service, the certainty associated with the average rating and the initial expectation of the service. Then, all the attribute trust is aggregated through a logic-based algorithm [139].

As everything can be a service in cloud environments, Noor and Sheng propose a concept named "Trust-as-a-Service" [118], and carry out a range of studies for trust management of cloud services [117, 120]. The trust result of every service is evaluated based on the aggregation of cloud consumers' feedback. They propose an approach based on majority consensus and feedback density to evaluate consumers' credibility of providing truthful assessments, so that the consumers' feedback can be weighted by their credibility. Furthermore, they propose an approach in [119] for the detection of collusion attacks and Sybil attacks in trust assessment.

In [47], Ghosh *et al.* propose a framework named "SelCSP" to compute the interaction risk of cloud providers for cloud consumers. The risk computation is based on the aggregation of the trustworthiness evaluation of providers through consumers' feedback or direct experiences and the competence evaluation of providers through the transparency of SLA guarantees. Finally, the risk levels of a cloud provider under different contexts can be estimated for a particular consumer.

In [105], Marudhadevi *et al.* propose a trust model for consumers selecting trustworthy cloud services based on previously monitored performance data and consumers' direct experiences. In this model, the trust degree of a cloud service is computed in two steps: 1) before signing a SLA, previous QoS data and consumers' feedback of the service are collected to compute the first trust degree (L1TD) for a consumer to judge whether to sign the SLA; and 2) if the degree is over a threshold, the consumer agrees to consume the service and provide his/her feedback for this service. Then based on the consumer's feedback, a new trust degree (L2TD) is computed through Bayesian inference to determine whether to continue using the service.

In [167], Wang and Wu propose a trust model based on a six-direction trust coordinate. The six directions of the coordinate consists of three groups of trust evaluations. The first one is the subjective-objective trust, where subjective trust refers to the trust evaluation based on users' subjective judgments; and objective trust is evaluated through quantitative performance measurements. The second group is the direct-indirect trust, where direct trust is extracted from a user' direct experience; and indirect trust comes from other users' recommendations. The last group is the inflow-outflow trust, where inflow trust and outflow trust respectively represent a service's received trust from the public and given trust to other services. Then a geometric algorithm is applied based on the coordinate for finding the most trustworthy service (finding the minimum coverage polyhedron).

In [32], Ding *et al.* propose a model for trustworthiness evaluation of cloud services, named "CSTrust", based on the aggregation of both customer satisfaction estimation and QoS prediction. Their work takes both qualitative attributes and quantitative attributes of cloud performance into account. The missing QoS values of quantitative attributes are predicted through collaborative filtering based on the performance records of other similar services. And cloud consumers' missing ratings for qualitative attributes can also be estimated through collaborative filtering. The results of objective prediction and subjective estimation are combined and weighted by customers' preference to compute the final service trustworthiness.

In [130], Qu and Buyya propose a hierarchy-based IaaS cloud selection approach. The novelty of the proposed approach is to determine users' requirements through a rule-based fuzzy inference system. In this work, the users' requirements can be expressed by various forms, e.g., numerical requirements and linguistic requirements. The inference system can take all these requirements as input, and then output the overall trust of every alternative service.

In general web service environments, Galizia *et al.* propose a trust-based selection approach for semantic web services [43]. The selection process is based on the Web Service Modeling Ontology [38], and modeled as a classification problem. It is assumed that every web service or user has a trust profile specifying its trust guarantee and requirements, which are determined by expert opinions. The selection result is a group of web services whose trust profiles match a user's trust profile.

Malik and Bouguettaya carry out a range of studies of credible trust establishment in service-oriented environments [99, 100, 101]. In their studies, a service consumer is allowed to share the ratings of the performance quality of services. In order to accurately assess service QoS, the rater's credibility is evaluated by comparing his/her ratings to majority consensus and the historical reputations of services. Thus, biased ratings or even malicious concluded ratings can be identified and set very low credibility. Then the credibility is used as weights to compute the new reputations of services.

In [163], Vu *et al.* propose a web service selection approach based on both service consumers' feedback and the values of the QoS attributes monitored by some trusted third parties. The feedback from the trusted agents is assumed to be fully trusted in the proposed work. According to the agents' feedback, the credibility of users is evaluated through comparing their feedback to the trusted feedback. In some cases, the credibility of a user cannot be evaluated directly since the user is isolated from any trusted agent. In such a case, the trustworthiness of the user's feedback is evaluated through the k-mean clustering algorithm for finding similar users. The final score of an alternative service is computed through the weighted feedback of users according to their credibility.

2.3.4 Description-based Selection

The studies of description-based selection are motivated by the problem of automatic cloud service selection without human labor. In this scenario, service providers are usually allowed to advertise their services using a machine-readable language, e.g., OWL [144] and WSML [82]. In the meantime, service requestors' preferences and requirements are also expressed via a broker using the same language. Then an ontology-based matching algorithm is employed to automatically find the qualified services according to the structured service definitions and requirements. In some cases, service real-time QoS information can also be gathered via autonomous agents for dynamic service selection.

In [177], Wittern *et al* focus on expressing cloud services through extended feature modeling, which is applied to represent the commonalities and differences of cloud services. The features of a cloud service can be functional or non-functional. Through variability modeling, all configurations of a services are summarized for further selection. The cloud service selection process is carried out by finding the configurations which fulfill all the required objectives. Such a process can be considered as a filter to reduce the number of valid service configurations. The outcomes of the selection process can be used as an input for any further decision-making approaches of service selection. A prototype is demonstrated in [177] for selecting real-world cloud storage services.

In [190], Zhang *et al.* propose a declarative recommender system, named "CloudRecommender", for cloud infrastructure service selection. The aim of this system is to achieve automatic service selection without human involvement. To this end, a formal mapping is established based on ontology between consumers' requirements and service configurations. A prototype is presented to demonstrate the effectiveness of the proposed system by carrying out a service selection process of real-world infrastructure service providers, including Windows Azure, Amazon, GoGrid, RackSpace, etc.

In [140], Ruiz-Alvarez and Humphrey propose a XML schema for the descriptions of cloud storage services. Through such descriptions, consumers' requirements can be automatically matched to the targeted services without human effort on reading service definition documents. In addition, the proposed schema can also be applied to describe local clusters, and thus the costs and performance changes of service transfer from local storage to cloud storage can be estimated. A user case is proposed in [140] for describing the cloud storage services on Amazon and Windows Azure.

In [52], Goscinski and Brock propose a generic framework for cloud-based service publication, discovery and selection via Web Services Description Language (WSDL) [164]. Through this framework, service providers are allowed to publish their current state and characteristics of services via dynamic attributes defined in the WSDL documents. A prototype with a web-based user interface is introduced to validate the effectiveness of the proposed work.

In [115], Ngan and Kanagasabai propose a semantic cloud service discovery and selection system based on OWL-S which is an ontology for describing semantic web services [53]. In the proposed system, complex service constraints are first expressed through logic-based rules, and then input in a dynamic rule engine for semantic matchmaking based on real-time ontology population and reasoning. The matching outcome is a group of services with the corresponding matching levels. Then the alternative services can be ranked through a pre-defined scoring function.

In general service-oriented environments, Xu proposes a service discovery and selection model by combining the pre-defined QoS definitions and service reputations extracted from consumers' feedback [180]. The QoS information is provided by UDDI registries [164]. The reputation of a service is evaluated by a centralized reputation manager which is in charge of collecting feedback. The calculation of the reputation scores is designed to reflect the recent performance of services. To this end, consumers' feedback is decayed by time in the feedback aggregation. The most recent feedback has the most important weight. The final score of a service is computed by weighting both QoS attributes and the corresponding reputations according to a consumer's requirement.

In [108] and [107], Maximilien and Singh propose a multi-agent-based framework for web service selection according to a QoS ontology. The selection is based on dynamic assessment of non-functional attributes. The whole framework is driven by a multi-agent system, in which autonomous agents are used to represent service consumers and providers, and responsible for dynamic configuration of web services and sharing QoS information with other agents. The definition of services and consumers' requirements are expressed via a XML policy language. In addition, the dependency between QoS attributes is taken into account, e.g., high quality performance causes high price.

Table 2.2 illustrates the main works in the literature of cloud service selection in terms of the proposed generic procedure.

Categories	Approaches
Defining purpose of cloud service selection	
Selecting a single service	[54, 118, 117, 105, 167, 169, 32, 43, 149, 150, 99, 100, 101, 162, 159, 110, 111, 138, 145, 143, 46, 45, 168] [25, 152, 130, 79, 80, 142, 137, 49, 113, 66, 131, 134, 133, 136, 135, 7, 161, 71, 8, 140, 52, 108, 107, 180]
Selecting a group of composed services	[58, 57, 170, 22, 90, 77, 129, 128, 89, 20, 60, 34, 94, 178, 104, 181]
Identifying roles of service selection participants	
End consumers	[54, 118, 117, 105, 167, 32, 170, 169, 43, 143, 131, 134, 133, 136, 135, 110, 111, 138, 25, 145, 159, 142, 162, 168] [152, 130, 149, 150, 49, 71, 7, 113, 137, 66, 161, 181, 99, 104, 100, 101, 178, 8, 20, 140, 52, 180]
Application cloud providers	[22, 79, 80, 90, 77, 89, 60, 34, 94]
Infrastructure cloud providers	[58, 57, 129, 128]
Service criterion modeling	
Service measurement index	[108, 107, 79, 49, 71, 143, 130, 142, 138, 91, 46, 45, 30]
Concrete attribute modeling	[104, 145, 44, 28, 191, 44, 28, 191, 62, 85]
User preference evaluation	[77, 113, 149, 150, 25, 195, 194, 134, 135]
Service criterion evaluation	
End-user-side QoS evaluation	[54, 118, 117, 119, 120, 47, 105, 167, 32, 99, 100, 101, 77, 131, 134, 133, 136, 135] [149, 150, 168, 49, 143, 46, 45, 152, 130, 108, 107, 180]
Service-side QoS evaluation	[108, 107, 169, 90, 34, 94, 137, 161, 162]
QoS description	[47, 43, 8, 170, 151, 129, 128, 181, 104, 20, 89, 22, 178, 79] [60, 80, 177, 190, 140, 52, 108, 107, 180, 58, 57]
Service selection execution	
Trust-based selection	[54, 118, 117, 47, 47, 105, 167, 32, 43, 8, 130]
MCDM-based selection	[79, 80, 137, 161, 66, 71, 25, 7, 110, 131, 134, 133, 136, 135, 111] [143, 113, 142, 145, 46, 159, 45, 138, 149, 150, 152, 162, 168]
Optimization-based selection	[58, 57, 170, 169, 22, 90, 77, 129, 128, 89, 20, 181, 151, 60, 34, 94, 178, 104]
Description-based selection	[43, 177, 190, 140, 52, 108, 107, 180]

Table 2.2: Summary of Cloud Service Selection

2.4 Related Techniques in Incentive Mechanism Design

Incentive mechanisms are applied under various contexts, e.g., e-marketplace, reputation systems, crowdsourcing, etc. In general, an incentive mechanism is typically designed for the purpose of eliciting the cooperation of participants who are rational and self-interested [12], i.e., every participant is motivated to maximize their own payoffs. Game Theory [116] is the fundamental for designing an effective incentive mechanism, under which an equilibrium can be achieved among all participants for some pre-defined goals, e.g., eliciting truthful assessments.

In our studies of cloud service selection, an uncertain compatible incentive mechanism is proposed to motivate cloud users providing continual and truthful assessments. The proposed mechanism can not only effectively improve the accuracy of cloud service selection since cloud users would try to provide less biased or noisy assessment under the mechanism, but also benefit the evaluation of the dynamic performance of cloud services.

In the literature, incentive mechanisms for eliciting truthful information are usually modeled in a seller-buyer scenario, where speaking the truth is an equilibrium for buyers. Suppose there are a group of sellers and a group of buyers. Buyers publish their satisfaction on the goods or services after they have transactions with sellers. The buyers who are considered to provide honest feedback will be rewarded, and the feedback will truthfully reflect sellers' reputation of offering satisfactory goods or services. According to the applied techniques, those mechanisms can generally be classified into two types: peer-prediction based approaches and reputation-based approaches.

2.4.1 Peer-prediction based Approaches

Miller *et al*. [112] propose the pioneering "Peer-Prediction" method for eliciting truthful feedback. In their work, every user can obtain monetary payment from an authorized center. The amount of payment depends on how well a user can predict the signal from some other user (called a reference user) based on its own signal. Their work is feasible based on several common knowledge assumptions, e.g., product type distributions and conditional distributions of signals. However, there is a drawback in Miller *et al.*'s work, i.e., there may exist lying equilibria that can bring higher expected payoffs than the truthful equilibrium [112].

To overcome this drawback, Jurca and Faltings [74, 75] propose a collusion-resistant feedback payment approach, in which several reference reports are applied in the scoring rules instead of the one-reference-report scheme in the prior work. They prove that speaking the truth is the unique equilibrium if at least three reports are used.

In the later studies, Witkowski [175] points out that the quality of goods or services provided by sellers is assumed fixed in prior works. However, in many real-world situations, the quality is inherently dynamic. Thus, he proposes a payment mechanism based on the hidden Markov setting to deal with such dynamics. It is worth noting that all these peer-prediction-based incentive mechanisms make strong common knowledge assumptions. To lift these assumptions, Witkowski and Parkes [176] propose peer prediction without a common prior. Their mechanism allows participants to adopt subjective and private priors instead of a common prior by asking a participant to offer two reports (one before the transaction and one afterwards), and their approach is proved to provide strict incentives for truthful reports. Compared to the peer-prediction-based approaches, our proposed incentive mechanism needs fewer knowledge assumptions and no extra belief report submission.

In [72], Jurca *et al.* propose an incentive compatible reputation mechanism. In this work, before two agents A and B have a transaction, they need to buy the reputation information of each other from a broker agent named *R-agent* to determine whether to have the transaction. If the transaction is completed, the two agents can sell the reputation information back to the same R-agent, and agent A can get paid only if the rating A submits for agent B is the same as the following rating for B submitted by another agent.

It can be concluded that the incentive mechanism introduced above involves a/an payment/award function, through which instant payment/award for a participant is calculated according to its feedback and others' feedback.

2.4.2 Reputation based Approaches

Some incentive mechanisms focus on evaluating participants' reputations on how truthfully they provide assessments or do something they have committed to. And the reputation would influence a participant's future opportunities of obtaining profits.

Jurca and Faltings [73] propose an incentive-compatible reputation mechanism, which allows sellers to "confess" when they did not provide the goods or services as those they have committed. The proposed mechanism focuses on dealing with opportunistic behavior where a seller builds an good reputation by providing satisfactory services or goods first, and then starts to cheat buyers later. The proposed mechanism is based on fact that a seller trends not to cheat a buyer who has a good reputation on telling the truth since the cost suffered from the negative report submitted by such a buyer would exceed the benefit from cheating. Both the seller and the buyer are asked to submit reports about the quality of the provided services or goods. If they cooperate (i.e., providing the same reports), the reputations of both of them increase, otherwise decrease. Due to such a confession, a seller can prevent further losses for his/her cheating, which give sellers incentives to speak the truth.

Papaioannou and Stamoulis [124], propose a reputation-based incentive mechanism in a peer-to-peer system to motivate peers for truthful reporting. In their work, a non-credibility metric is designed for controlling a peer's punishment of having disagreed transaction feedback with other peers. After each transaction, the peers involved in the transaction would be asked to provide reports on whether the transaction is satisfactory. If the submitted reports are the same, the non-credibility values of all the peers decrease as rewards, otherwise increase as punishments. Furthermore, uncooperative peers would be forbidden to participate in further transactions for a period of time which is exponentially related to their non-credibility. This motivates peers to always give truthful reporting in order to avoid punishments.

Zhang *et al.* [189] propose a trust-based incentive mechanism, which is an extension of their prior work [188], in a reverse auction scenario. In this mechanism, a buyer computes personalized reputations of other buyers based on the aggregation of both public and private reputation evaluations. A buyer is considered trustworthy if many other buyers have similar reports to the buyers, and thus believe he/she has a good reputation. On the other hand, sellers compute the reputations of buyers. A buyer having a good reputation would be rewarded by sellers since the report submitted by such a buyer is considered more valuable for sellers to build good reputations. Furthermore, a seller whose reputation is below a threshold is forbidden to participate in future auctions and therefore suffers a loss. Thus, the proposed mechanism can benefit both honest buyers and sellers in the marketplace.

2.4.3 Incentive Mechanism Studies in Crowdsourcing

In addition to the above approaches, some recent studies of incentive mechanisms in crowdsourcing environments are proposed for eliciting effective contributions of workers. In general, all these incentive approaches are proposed for eliciting the cooperation of users, and therefore are related to our studies.

In [106], Mason and Watts study the relationship between financial incentives and working performance, and argue that increasing financial incentives could only bring more workers, but not a working quality improvement as expected. A similar conclusion can be found in DiPalantino and Vojnovic's work [33]. They argue that worker participation rates logarithmically increase with monetary rewards.

In [193], Zhang and van der Schaar focus on solving workers' *free-riding* problem and requesters' *false-reporting* problem. They designed optimal and sustainable incentive protocols based on social norms [76]. In the proposed protocol, a service requester must pay workers in advance no matter how well the workers work. The aim of this setting is to prevent requesters from falsely reporting working quality. On the other hand, a worker's working quality is judged by the requester. Because of the setting of ex-ante payments, the requester would always tell the truth about working quality since cheating cannot bring any profit. According to the requester's judgment, the workers' reputations are computed. If a worker's reputation is lower than a predefined threshold, which is determined by a range of protocol parameters (e.g., the probability of the requester falsely judging or a worker's confidence on obtaining further profits), the worker would be isolated from further transactions for a period of time. Thus, workers would always work well in order to obtain more profit.

In Zhang *et al.*'s further study [192], a generic rating protocol is proposed for any type of online communities. Through theoretical analysis, the proposed optimal protocol is proved to not only feasibly sustain cooperation among users but also solve the *whitewashing* problem [41].

2.5 Open Issues in Cloud Service Selection

According to our survey, we discuss some open issues which have not been sufficiently studied in the literature of cloud service selection.

- **Lack of wide-recognized standards of cloud performance:** due to the diversity of cloud services and cloud consumers' customized requirements, setting a common-acknowledged standard for identifying all cloud performance attributes and quantitatively describing these attributes is believed impossible in practice. However, lack of such a standard results that human-decision procedure have to be involved in cloud selection, and thus fully automatic service selection cannot be achieved in real-world situations.

  Most studies of cloud service selection focus on the selection based on general performance attributes without specifying what the attributes are and how to model them. Only a few studies proposed cloud performance taxonomy or ontology [146, 45, 143, 149, 79], but these studies have not be commonly recognized since cloud consumers have very different and customized views in cloud selection. So far, this is still a issue which cannot be effectively solved in the application level. If cloud performance can be standardized, the effectiveness and efficiency of cloud service selection can be greatly improved.

- **Lack of classification of cloud performance assessments:** in the literature, the different types of assessments of cloud performance are usually treated equally and would be typically taken as the input of a MCDM approach for service selection, which may introduce numerous noise into service selection since assessments given under quite different situations may be quite different. For example, from the perspective of service delivery, cloud assessments can be given from the service side or the end-user side. There may exist a large discrepancy of cloud performance between both side. In addition, from the perspective of who offering assessments, cloud assessments can be classified into subjective assessments and objective assessments. Either type of assessments should be differently treated. That is because, objective assessments may not be intuitive for ordinary consumers, who may feel confused when facing numerous quantitative information. On the other hand, although subjective assessments are easy to get understanding, they may contain bias and malicious assessments, and thus cannot be fully trusted. In the current studies of cloud service selection, the classification of cloud assessments before carrying out service selection has not been paid enough attention.

- **Lack of evaluation of assessment credibility according to assessment classification:** whatever assessments are used in cloud service selection, the credibility of assessments should be carefully evaluated in order to make the selection more accurate. Due to the different types of cloud assessments introduced in the last paragraph, the ways of computing assessment credibility should be customized according to assessment types. For example, objective assessments should be evaluated according to the contexts under which the assessment are given. And subjective assessment evaluation should focus on filtering out unreasonable ones using bias or malevolence detection techniques. In a nutshell, different types of assessments should be evaluated in different ways according to their features. This issue has not been sufficiently studied in the literature.

- **Lack of evaluation of long-term performance of cloud services:** as introduced in Section 1.1, dynamic performance is a main feature of cloud computing. A potential cloud consumer desires to know if the selected service can maintain its service level over long time. To this end, cloud performance evaluation should consider the long-term performance of services. The variation of service performance should be monitored via long-term assessments from ordinary consumers or professional testing parties. In addition, the credibility of assessments should also be evaluated over long time in order to judge how much the parties who offer assessments can be trusted. In the literature, most of studies apply time-decay functions to aggregate cloud assessments over time. Through this way, the variation of cloud performance cannot be demonstrated to consumers. And thus the overall performance of cloud services cannot be reflected.

- **Lack of incentives for eliciting long-term cloud assessments:** since long-term assessments play a very important role in evaluating the dynamic performance of cloud services, there should be a way to motivate cloud consumers to regularly provide continual and truthful assessments. This can be considered as the active way to improve assessment reliability, compared to the traditional way in the literature, i.e., passively evaluate assessment credibility. However, to the best of my knowledge, there is no such an incentive mechanism proposed in the literature.

In this thesis, we propose some solutions for those issues discussed above. The experimental results and theoretical analysis prove the feasibility of our solutions.

2.6 Conclusion

In this chapter, we first introduce the background knowledge of cloud computing, and then propose a generic procedure of cloud service selection, consisting of five steps: defining the purpose of cloud service selection, service criterion modeling, identifying the roles of service selection participants, service criterion evaluation and service selection execution. The proposed procedure aims to cover all the related studies in the literature, and thus identify popular research focus areas as well as open issues. Then, we classify the related models and approaches of cloud service selection based on the generic procedure, and summarize the main techniques of cloud service selection in the literature. Then, the related techniques and studies of incentive mechanism design are introduced. Finally, some open issues in cloud service selection are discussed.

Chapter 3

Cloud Selection based on Subjective Assessments and Objective Assessments

Cloud computing has been attracting huge attention in recent years. Because of the outstanding advantages of cloud computing, e.g., flexibility and low cost, more and more individuals and organizations have started to consume cloud services. It should be noted that the emergence of cloud services also comes with new challenges. One big challenge is how to evaluate the performance of cloud services. To this end, many cloud performance monitoring, testing and comparison approaches [45] [85] [137] have been proposed. The usual way for cloud service evaluation is to compare the performance differences between similar cloud services. Such a comparison is usually based on the results of a predesigned set of benchmark tools [85, 84]. As cloud services are highly virtualized, the benchmark tools for traditional computation performance measurement can be appropriately applied in cloud environments. By combining these benchmark tools according to cloud features, many metrics can be quantitatively measured (e.g., the speed of CPU, memory read/write and storage, service response time and throughput).

Nevertheless, the benchmark testing results usually may not reflect the real performance of cloud services for ordinary cloud consumers. This is because the testing environment is usually not the same as that of ordinary consumers' daily work, and a variety of real tasks currently executed in a cloud may not be perfectly simulated by a limited number of tests. In addition, these benchmark tests are usually spot-check tests. It is hard to carry out continuous testing because such tests might lead to costing no less than that of consuming a real cloud service. Furthermore, some crucial but qualitative aspects of cloud services can hardly be tested through such objective and quantitative measurement. For example, considering a company working on processing a large amount of sensitive customer data, the security and privacy of the data have a crucial impact on the company's survival. If the company plans to move their work into clouds in order to reduce the optional costs, it must choose a cloud provider which has a very good reputation on data security and privacy. In addition, as the company is not a professional IT company, good and comprehensive after-sales services are highly desired. Moreover, due to the sensitivity of the company's data, a variety of encryption approaches are frequently applied in daily work. Hence, the speed of data encryption and decryption is a big concern for the company. In this example, in addition to the typical performance of a cloud service (e.g., CPU, memory, response time and costs) that can be quantitatively tested by common benchmark testing, the company needs to carefully consider the data privacy and security and the quality of after-sales services, which can hardly be quantified. In addition, the benchmark tests of the cryptographic calculation speed of a cloud service may need to be specifically designed according to the company's requirement. All the issues mentioned above may cause great difficulty for the company in cloud service selection.

In this chapter, we propose a novel model based on the aggregation of the feedback from cloud consumers and the objective performance measurement from a trusted third party's testing. A framework supporting the proposed model is first presented in Section 3.1. Then, in Section 3.2, we present the details of our cloud service selection model that evaluates the performance of cloud services by aggregating all the subjective assessments and the objective assessments through a fuzzy simple additive weighting system [26]. It should be noted that cloud consumers' subjective assessments may be biased and inaccurate since they are usually not professional IT staff and even some of them may be malicious users. Hence, in our model, cloud consumers' subjective assessments and the third party's objective assessment are compared, so that unreasonable subjective assessments can be filtered out before aggregation. This makes our approach more accurate and effective. Finally, a case study is presented in Section 3.3 to illustrate the advantages of our model.

3.1 The Cloud Selection Framework

In this section, we present our framework of cloud service selection based on both the feedback from cloud users and the objective performance benchmark testing from a trusted third party. Figure 3.1 illustrates our framework, which consists of four components, namely, (1) *cloud selection service*, (2) *benchmark testing service*, (3) *user feedback management service*, and (4) *assessment aggregation service*, where *cloud selection service* is in the higher layer of our framework, and others are in the lower layer.

3.1.1 Cloud Selection Service

The *cloud selection service* is in charge of accepting and undertaking the preliminarily processing of requests for cloud service selection from potential cloud consumers. In addition, it issues the requests for the services from the lower layer components. When a potential cloud consumer submits a request for selecting the most suitable cloud service, the *cloud selection service* firstly chooses the cloud services which can meet all the objectives and quantitative requirements (e.g., the type of services, the specification of virtual machines and costs) of the potential user from a candidate list of cloud services. Then, according to the user's further requirements, it sends requests to the *benchmark testing service* and the *user feedback management service* for accessing the related records of alternative clouds. These records are then sent to the *assessment aggregation service*. By aggregating these records through our proposed model (see

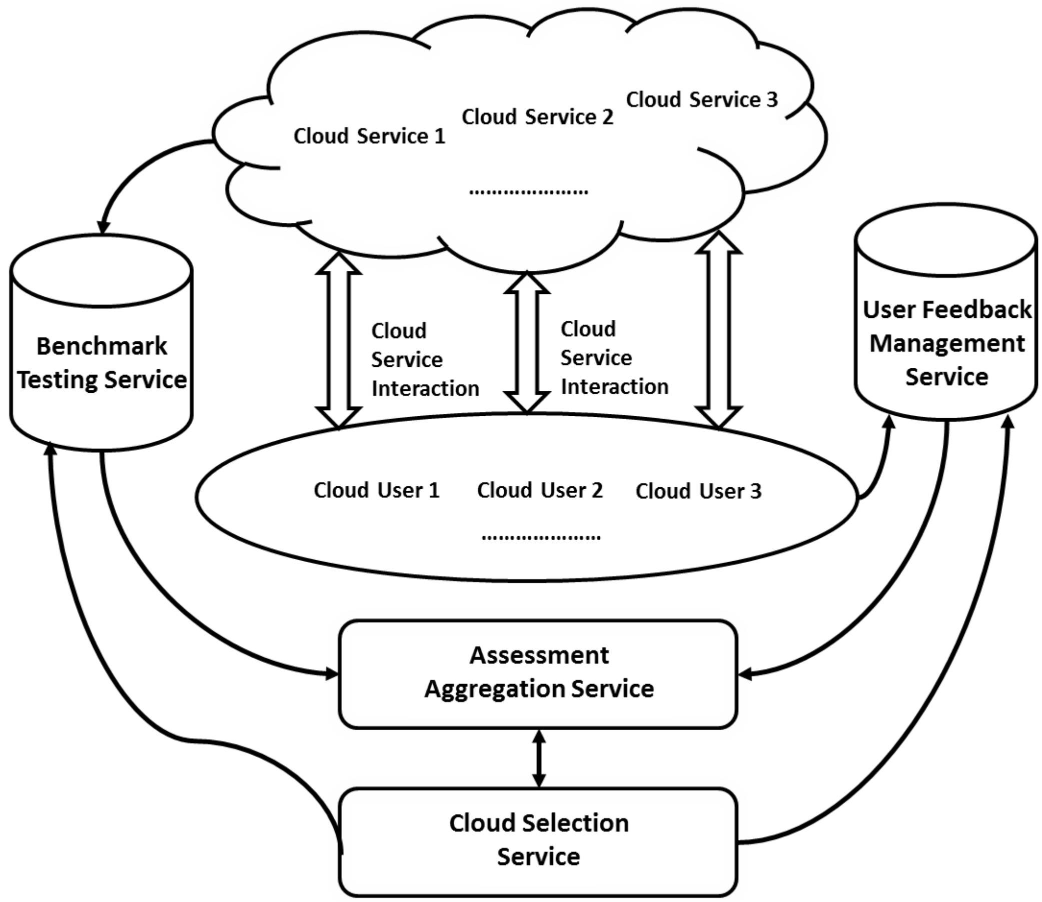


Figure 3.1: Our Proposed Framework for Cloud Service Selection

details in Section 3.2), the *assessment aggregation service* returns the final score of each alternative cloud to the *cloud selection service*. All these scores are shown to the potential user for cloud selection.

3.1.2 Benchmark Testing Service

The *benchmark testing service* is provided by a trusted third party which designs a variety of testing scenarios for the common performance aspects of a cloud service (e.g., availability, elasticity, service response time, and cost per task) by standard benchmark suites. In addition, some specific tests can be designed and run according to potential cloud consumers' needs, such as testing the speed of cryptographic calculations. The common performance tests can be executed in the spot-check form over time for a long period. The specific tests can run in a continuous but short period according to potential users' requirements. Each tested performance aspect of a cloud service can be considered as the *objective attribute* of the cloud service. All these objective attributes are expressed in quantified forms (e.g., 90% for availability, $200ms$ for response time or 47.5 benchmark scores for CPU performance). All the attribute values are recorded and maintained for the requests from the *cloud selection service*.

3.1.3 User Feedback Management Service

The *user feedback management service* is used to collect and manage the feedback from the users who are consuming cloud services. For each performance aspect of a cloud service, a user gives his/her subjective assessment according to his/her intuitive feelings. Each aspect that users assess can be considered as the *subjective attribute* of the cloud service. These subjective attributes are expressed by linguistic variables (e.g., "good", "fair" and "poor").

Note that some subjective attributes and some objective attributes can represent the same performance aspect of a cloud service. For example, the response time of a cloud service can be accurately calculated by benchmark testing under different circumstances. By analyzing the testing results, an objective assessment of the service response time can be given. Meanwhile, a user of this cloud can also give subjective assessments of response time by feeling how long the cloud responds to his/her requests in daily work. Figure 3.2 introduces an example to explain the relationship between these attributes. In our framework, such attributes (e.g., response time which belongs to both subjective attributes and objective attributes) are named as *associated attributes*. In Figure 3.2, for a cloud service, we assume there are s subjective attributes, o objective attributes and u associated attributes for a cloud service ($u \leqslant s, u \leqslant o$), where *privacy, after-sales services, availability* and *response time*

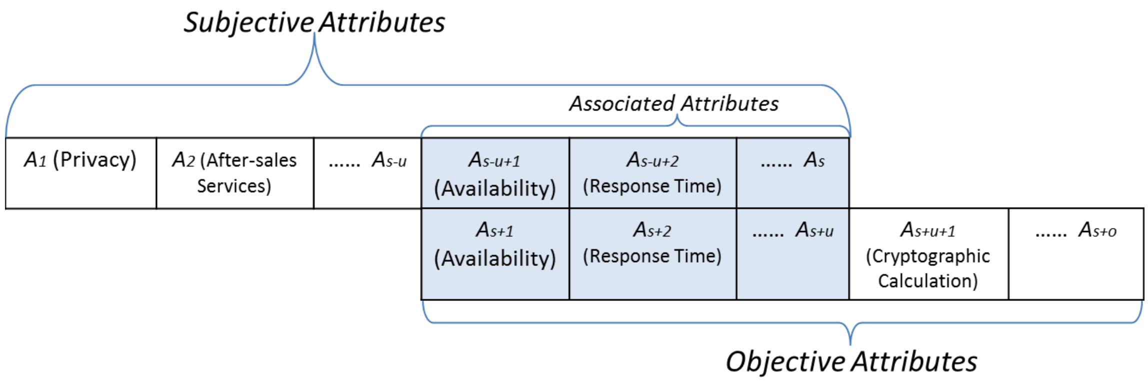


Figure 3.2: The Relationship of Subjective Attributes and the Objective Attributes

are its *subjective attributes* extracted from users' feedback. On the other hand, it also has *availability, response time* and *cryptographic calculation* as its *objective attributes* extracted from a third party's benchmark testing. Thus, *availability* and *response time* are considered as the *associated attributes*. The *associated attributes* from subjective assessment are called *subjective associated attributes*, and those from objective assessment are called *objective associated attributes*. Therefore, each *subjective associated attribute* corresponds to an *objective associated attribute*.

Furthermore, in some cases, the users' subjective assessment of a cloud service may not be absolutely subjective. For those users with an IT background, some simple testing tools or status checking commands (e.g., Command *xentop* for Xen Hypervisor[1]) can be used to help them make an assessment. In such a situation, we consider that the assessment still belongs to subjective assessment since the results from such cloud users' testing are usually incomprehensive and inaccurate without scientific and statistical analysis.

3.1.4 Assessment Aggregation Service

The *assessment aggregation service* is in charge of aggregating the values of the subjective attributes from the *user feedback management service* and the objective at-

[1] xen.org

tributes from the *benchmark testing service*, and computing the final score for each alternative cloud service according to the importance weights that are set by a potential cloud user in the form of linguistic variables. In the company example introduced before, suppose that the company is quite concerned about the issues of user data's security and privacy and hopes to receive excellent after-sales service. Therefore, the importance weights for *security* and *privacy* as the attribute of a cloud service can be set as "very high", and the weight for *after-sales service* can be set as "high". Likewise, all the other subjective attributes and objective attributes are given such importance weights. By using these weights, the potential cloud user (i.e., the company) can also determine whether to put more trust on subjective assessment or objective assessment, so that the final score based on aggregating all these attributes can comprehensively reflect the various needs of the potential cloud user (i.e., the company).

3.2 The Cloud Service Selection Model

In this section, we propose a novel cloud service selection model which is based on comparing and aggregating the subjective attributes and the objective attributes of a cloud service. In our model, for the sake of simplicity, we assume there are sufficient users offering their subjective assessments, and there is only one third party offering objective assessment for a cloud service. That is, there is only one set of values for objective attributes and there are many sets of values for subjective attributes for one cloud service. In addition, the third party is assumed to be a trusted third party which offers honest results of the objective performance testing. The trustworthiness of the third party and the situation with multiple third parties will be considered in Chapter 6. Hence, in our model, we take objective assessment as a benchmark to evaluate the accuracy of subjective assessments, since objective assessment is usually more accurate to reflect the real performance of a cloud service than subject feeling.

In order to compare and aggregate subjective attributes and objective attributes, the values of these attributes should be normalized. Here, we apply a fuzzy simple additive weighting system (FSAWS) proposed by Chou *et al.* [26] to convert all the attribute values into ratings. Firstly, after gathering both the values of the subjective attributes and the objective attributes of an alternative cloud service, the values of the subjective attributes are converted into ratings through a mapping from linguistic variables to fuzzy numbers. Secondly, the values of the objective attributes are also converted into the ratings by comparing the values of the same objective attributes in all the alternative cloud services. Thirdly, as introduced above, for an alternative cloud service there are many sets of values for the subjective attributes from different users' feedback and only one set of values for the objective attributes from the third party's testing. We compute the Euclidean distance [31] between the set of values for the objective associated attributes and each set of the values for the subjective associated attributes through the corresponding normalized ratings. The set of values for the subjective attributes with the Euclidean distance over a threshold (e.g., 80% of the maximum distance) will be considered unreasonable and eliminated from the following aggregation of all the attributes. After that, the importance weight for each attribute is computed according to the potential cloud user's preference. Finally, a score is calculated for each alternative cloud service. The higher the score, the better the overall performance of an alternative cloud service.

In real world situations, subjective assessment for cloud services and importance weight for each performance attribute are usually represented in the form of linguistic variables (e.g., "good" and "bad"). In order to deal with the inherent uncertainty of human languages, we apply a fuzzy simple additive weighting system in our model. Through this system, linguistic variables can be represented by fuzzy numbers for their fuzziness. And quantitative terms can also be represented in fuzzy number form. Hence, by using this system, our model can effectively normalize and aggregate all different types of subjective attributes and objective attributes in real world situations. Before presenting the details of our model, some basic knowledge of FSAWS will be introduced.

3.2.1 Fuzzy Simple Additive Weighting System

The fuzzy simple additive weighting system [26] is originally proposed for solving the facility location selection problem which is a multiple attributes decision-making problem under homo/heterogeneous group decision-making environments. In this system, the decision makers are grouped together to determine the most suitable facility location. The evaluation of an alternative location depends on a variety of subjective attributes and objective attributes. The importance weight and the rating value for each attribute are represented by trapezoidal fuzzy numbers. An aggregated fuzzy score is calculated for each alternative location through some fuzzy set operations. Then, each fuzzy score is defuzzified to a crisp number as the final score of each alternative location.

To explain how this system works, some basic knowledge of fuzzy numbers is introduced in [26].

1. Trapezoidal Fuzzy Number:

$\widetilde{A} = (a, b, c, d)$ is a fuzzy set on $\mathbb{R}$, where $a \leqslant b \leqslant c \leqslant d$ are real numbers. If

$$\mu_{\widetilde{A}}(x) = \begin{cases} \dfrac{x-a}{b-a}, & a \leqslant x \leqslant b \\ 1, & b \leqslant x \leqslant c \\ \dfrac{x-d}{c-d}, & c \leqslant x \leqslant d \\ 0, & \text{otherwise,} \end{cases}$$

then $\widetilde{A} = (a, b, c, d)$ is called a trapezoidal fuzzy number, where $\mu_{\widetilde{A}}(x)$ is its membership function. The most probable value of the evaluation data is represented in the interval $[b, c]$. The intervals $[a, b]$ and $[c, d]$ show the fuzziness of the evaluation data. For example, "good" can be represented by $(5, 7, 7, 10)$, and "very good" can be represented by $(7, 10, 10, 10)$. In addition, some quantitative terms can also be represented by trapezoidal fuzzy numbers. For example, "equal to 50" can be represented by $(50, 50, 50, 50)$, and "approximately equal to 700" can be represented by

$(690, 700, 700, 710)$.

2. Operations of Trapezoidal Fuzzy Numbers:

Given two trapezoidal fuzzy numbers $\widetilde{A} = (a, b, c, d)$ and $\widetilde{B} = (e, f, g, h)$, and a real number k, some operations are defined as follows:

Addition:

$\widetilde{A} \oplus \widetilde{B} = (a + e, b + f, c + g, d + h), a \geqslant 0, e \geqslant 0.$

Multiplication:

$\widetilde{A} \otimes \widetilde{B} = (ae, bf, cg, dh), a \geqslant 0, e \geqslant 0;$

$k \otimes \widetilde{A} = (ka, kb, kc, kd), a \geqslant 0, k \geqslant 0;$

$\widetilde{A} \otimes k = (ka, kb, kc, kd), a \geqslant 0, k \geqslant 0.$

Division:

$\widetilde{A}/\widetilde{B} = (\frac{a}{h}, \frac{b}{g}, \frac{c}{f}, \frac{d}{e}), a \geqslant 0, e > 0;$

$k/\widetilde{A} = (\frac{k}{d}, \frac{k}{c}, \frac{k}{b}, \frac{k}{a}), a > 0, k \geqslant 0;$

$\widetilde{A}/k = (\frac{a}{k}, \frac{b}{k}, \frac{c}{k}, \frac{d}{k}), a \geqslant 0, k > 0.$

Commutative operations:

$\widetilde{A} \oplus \widetilde{B} = \widetilde{B} \oplus \widetilde{A}$, $a \geqslant 0, e \geqslant 0;$

$\widetilde{A} \otimes \widetilde{B} = \widetilde{B} \otimes \widetilde{A}$, $k \otimes \widetilde{A} = \widetilde{A} \otimes k, k \geqslant 0, a \geqslant 0, e \geqslant 0.$

3. Defuzzification:

Taking a trapezoidal fuzzy number $\widetilde{A} = (a, b, c, d)$ as input, the defuzzified output is a crisp number defined by computing the signed distance of $\widetilde{A}$:

$$d(\widetilde{A}) = \frac{1}{4}(a + b + c + d). \tag{3.1}$$

3.2.2 The Cloud Selection Approach

In this section, we present the details of our cloud service selection model. We modify the fuzzy simple additive weighting system [26] in order to fit our targeted problem. In addition, in our model, unreasonable subjective assessments can be filtered out before the aggregation of all the assessments. Therefore, the aggregated score of each alternative cloud service can more accurately reflect the overall performance of a cloud service with less noise.

Assume that a potential cloud user submits its request to the *cloud selection service* for finding the most suitable cloud service meeting all the user's requirements. After the preliminary selection according to the functional requirements, suppose that there are m clouds left as the alternative clouds denoted by $\{C_j\}$, where $j = 1, \cdots, m$. The final score of each alternative cloud is computed based on s subjective attributes extracted from cloud users' feedback and o objective attributes extracted from the benchmark testing of a trusted third party, where there are u ($u \leqslant s, u \leqslant o$) associated attributes. Thus, all the subjective attributes and the objective attributes are denoted as $\{A_i\}$, where $i = 1, 2, \cdots, s + o$. $\{A_i\}$ ($i = 1, \cdots, s$) denotes the subjective attributes, where $\{A_i\}$ ($i = s - u + 1, \cdots, s$) denotes the subjective associated attributes; $\{A_i\}$ ($i = s + 1, \cdots, s + o$) denotes the objective attributes, where $\{A_i\}$ ($i = s + 1, \cdots, s + u$) denotes the objective associated attributes. The corresponding objective associated attribute of the subjective associated attribute A_i is A_{i+u} for each $i = s - u + 1, \cdots, s$.

For an alternative cloud C_j, the *user feedback management service* returns n feedbacks given by n cloud users. Each feedback includes s values in the form of linguistic variables corresponding to the s subjective attributes. The feedbacks are denoted as $\{F_{jk}\}$, where $k = 1, \cdots, n$. On the other hand, the *benchmark testing service* returns one benchmark testing report for each alternative cloud. Each report includes o values in the form of quantitative terms corresponding to the o objective attributes. The reports are denoted as $\{T_j\}$, where $j = 1, \cdots, m$. m is the number of all the alternative cloud services.

It should be clarified that, for an alternative cloud, there are n feedbacks and only one testing report. That is, all the n feedbacks correspond to only one testing report since we assume there is only one trusted third party in our model. Here, for an alternative cloud, we connect the set of values for the o objective attributes from the only one testing report to the set of values for the s subjective attributes from each

Decision makers \ Attributes	1	2	$\cdots$	$s+o$
1	A_{1j1}	A_{2j1}	$\cdots$	$A_{(s+o)j1}$
2	A_{1j2}	A_{2j2}	$\cdots$	$A_{(s+o)j2}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n	A_{1jn}	A_{2jn}	$\cdots$	$A_{(s+o)jn}$

Table 3.1: Decision makers of an alternative cloud C_j

feedback. The final score of an alternative cloud is computed based on aggregating all the sets of values of the $s+o$ attributes. We define such a set of values of the $s+o$ attributes as a decision maker, denoted as DM_{jk} ($k = 1, \cdots, n$), for an alternative cloud C_j. Thus, the value of ith attribute of the jth alternative cloud from the kth decision maker can be denoted as A_{ijk}.

Table 3.1 illustrates the decision makers of an alternative cloud C_j. There are n decision makers for C_j. n is the number of user feedbacks for C_j. Note that, in Table 3.1, for each $i = s+1, \cdots, s+o$, all the A_{ijk} are the same for different decision makers DM_{jk} since they are all the values of the objective attributes from the only one testing report T_j. That is, for C_j, $A_{ij1} = A_{ij2} = \cdots = A_{ijn}$ for each $i = s+1, \cdots, s+o$.

The detailed procedure of our approach is shown below:

Step 1 (Converting the values of subjective attributes into ratings): By using a mapping from linguistic variables to fuzzy numbers, the linguistic values for the subjective attributes can be converted into fuzzy numbers. We use the mapping illustrated in Table 3.2 from [26], which is frequently employed in research of the multi-criteria decision-making problem for real world situations, such as in [93] and [92]. Each fuzzy number in Table 3.2 represents the fuzzy rating corresponding to the linguistic variable, denoted as $\widetilde{r}$. In addition, a crisp rating corresponding to each linguistic variable is computed by defuzzifying its fuzzy rating with the signed distance (Equation (3.1)), which is denoted as r. Hence, for a decision maker DM_{jk} of an alternative cloud C_j, $\widetilde{r}_{ijk}$ and r_{ijk} denote the fuzzy rating and the crisp rating for A_{ijk} respectively, where $i = 1, \cdots, s$.

Linguistic Variables	**Fuzzy Ratings** ($\widetilde{r}$)	**Crisp Ratings** (r)
Very poor (VP)	$(0, 0, 0, 20)$	5
Between very poor and poor (B.VP&P)	$(0, 0, 20, 40)$	15
Poor (P)	$(0, 20, 20, 40)$	20
Between poor and fair (B.P&F)	$(0, 20, 50, 70)$	35
Fair (F)	$(30, 50, 50, 70)$	50
Between fair and good (B.F&G)	$(30, 50, 80, 100)$	65
Good (G)	$(60, 80, 80, 100)$	80
Between good and very good (B.G&VG)	$(60, 80, 100, 100)$	85
Very good (VG)	$(80, 100, 100, 100)$	95

Table 3.2: Mapping from Linguistic Variables to Fuzzy Ratings and Crisp Ratings

Step 2 (Converting the values of objective attributes into ratings): As we have introduced in Section 3.1, the values in the form of quantitative terms can also be represented by fuzzy numbers. For any decision maker DM_{jk} of an alternative cloud C_j, the values of the objective attributes $\{A_{ijk}\}$ ($i = s+1, \cdots, s+o$) are represented in the form of fuzzy numbers. Then, these fuzzy numbers are converted into fuzzy ratings by comparing the values of the same objective attribute in all the alternative clouds. Let $\widetilde{x}_{ijk} = (a_{ijk}, b_{ijk}, c_{ijk}, d_{ijk})$ ($i = s+1, \cdots, s+o$) be the fuzzy numbers of the objective attribute values of the cloud C_j for any DM_{jk}. The fuzzy rating of each objective attribute value is computed as follows:

$$\begin{aligned} &\widetilde{r}_{ijk} = (\widetilde{x}_{ijk} / \max_{j}(d_{ijk})) \otimes 100, \text{where} \\ &i = s+1, \cdots, s+o, j = 1, \cdots, m, k = 1, \cdots, n. \end{aligned} \tag{3.2}$$

$$\begin{aligned} &\widetilde{r}_{ijk} = (\min_{j}(a_{ijk}) / \widetilde{x}_{ijk}) \otimes 100, \text{where} \\ &i = s+1, \cdots, s+o, j = 1, \cdots, m, k = 1, \cdots, n. \end{aligned} \tag{3.3}$$

Equation (3.2) is for the situation that the larger objective attribute value is the better (e.g., benchmark score for CPU). And Equation (3.3) is for the situation that the smaller objective attribute value is the better (e.g., response time).

Step 3 (Filtering unreasonable subjective assessments): So far, all the values in Table 3.1 have been converted into the form of fuzzy ratings. For each decision maker

DM_{jk} $(k = 1, \cdots, n)$ of an alternative cloud C_j, the Euclidean distance between the ratings of the corresponding subjective associated attributes and the objective associated attributes is computed as follows:

$$\begin{aligned} ED_{jk} &= \sqrt{\sum_{i=s-u+1}^{s} (d(\widetilde{r}_{ijk}) - d(\widetilde{r}_{(i+u)jk}))^2} \\ &= \sqrt{\sum_{i=s-u+1}^{s} (r_{ijk} - r_{(i+u)jk})^2} \quad . \end{aligned} \tag{3.4}$$

ED_{jk} represents the aggregated difference between the crisp ratings of the corresponding subjective associated attributes and the objective associated attributes of the decision maker DM_{jk} for the alternative cloud C_j. In our model, we take the objective assessment as the benchmark to filter out unreasonable subjective assessments. If the distance exceeds a threshold (e.g., 80% of the maximum Euclidean distance), the decision maker offering such values of the subjective attributes will be removed from the list of the decision makers for an alternative cloud.

Step 4 (Computing the importance weight for each attribute): According to the potential cloud user' requirement, an importance weight in the form of linguistic variables is given to each subjective or objective attribute, so that the user can determine the concern degree of each attribute. In addition, the user can also determine how much to trust the subjective or objective assessment through these importance weights. A fuzzy weight in the form of fuzzy numbers is given to each attribute, which is denoted as $\widetilde{W}_i$, where $i = 1, \cdots, s+o$. Table 3.3 [26] illustrates the mapping from linguistic variables to fuzzy weights. This mapping is also frequently used in prior studies for real world situations. Then, the fuzzy weights are defuzzified by computing their signed distances (Equation (3.1)). The crisp weight of the attribute A_i is denoted as W_i which is computed as follows:

$$W_i = \frac{d(\widetilde{W}_i)}{\sum_{i=1}^{s+o} d(\widetilde{W}_i)}, \text{where } \; i = 1, \cdots, s+o. \tag{3.5}$$

Linguistic Variables	**Fuzzy Weights**
Very low (VL)	$(0, 0, 0, 3)$
Low (L)	$(0, 3, 3, 5)$
Medium (M)	$(2, 5, 5, 8)$
High (H)	$(5, 7, 7, 10)$
Very High (VH)	$(7, 10, 10, 10)$

Table 3.3: Mapping from Linguistic Variables to Fuzzy Weights

Step 5 (Aggregating all attributes): Assume there are n' decision makers left for the cloud C_j after Step 3. A matrix $\widetilde{M_j}$ based on the fuzzy ratings of each attribute from different decision makers is constructed as follows:

$$\widetilde{M_j} = \begin{bmatrix} \widetilde{r}_{1j1} & \widetilde{r}_{2j1} & \cdots & \widetilde{r}_{(s+o)j1} \\ \widetilde{r}_{1j2} & \widetilde{r}_{2j2} & \cdots & \widetilde{r}_{(s+o)j2} \\ \cdots & \cdots & \cdots & \cdots \\ \widetilde{r}_{1jn'} & \widetilde{r}_{2jn'} & \cdots & \widetilde{r}_{(s+o)jn'} \end{bmatrix} .$$

According to the crisp weight of each attribute, the fuzzy scores of the cloud C_j from every decision maker DM_{jk} ($k = 1, \cdots, n'$) are computed as follows:

$$\widetilde{S_j} = \widetilde{M_j} \otimes \begin{bmatrix} W_1 \\ W_2 \\ \cdots \\ W_{s+o} \end{bmatrix} = \begin{bmatrix} \widetilde{f}_{j1} \\ \widetilde{f}_{j2} \\ \cdots \\ \widetilde{f}_{jn'} \end{bmatrix}, \tag{3.6}$$

where $\widetilde{f}_{jk}$ is the fuzzy score of C_j from DM_{jk} ($k = 1, \cdots, n'$). Here, the operation $\otimes$ is generalized to matrices in the standard way of matrix multiplication. Then, the final score of C_j is computed as follows:

$$\overline{S_j} = \frac{1}{n'}(\sum_{k=1}^{n'} d(\widetilde{f}_{jk})) \ . \tag{3.7}$$

Finally, according to the final scores, all the alternative cloud services are ranked for the selection of the potential cloud user.

3.3 A Case Study

In this section, a case study over the example introduced at the beginning of the chapter is presented to illustrate the effectiveness of our proposed model for cloud service selection.

Assume that the company which plans to consume a cloud service requests to select the most suitable cloud services based on seven attributes, including four subjective attributes and three objective attributes. The four subjective attributes are *privacy* (A_1), *after-sales service* (A_2), *availability* (A_3) and *service response time* (A_4); and the three objective attributes are *availability* (A_5), *service response time* (A_6) and *cryptographic calculation speed* (A_7), where *availability* and *service response time* are the associated attributes.

Step 1: Assume that there are four alternative cloud services left after preliminary selection. For each alternative cloud, there are ten feedbacks from the *user feedback management service* to assess the four subjective attributes in the form of linguistic variables. According to Table 3.2, all the linguistic variables are mapped into the fuzzy numbers which represent the fuzzy ratings for the subjective attributes.

Step 2: Every alternative cloud service has a performance report for the three objective attributes from the *benchmark testing service*. The value of the attribute *availability* is the average percentage of the times of successfully accessing a cloud service over the last twelve months. The value of the attribute *service response time* is the average of the service response time of every spot-check test over the last twelve months. The fuzziness of these values depends on the minimum value and the maximum value for each attribute recorded in the last twelve months. The speed of cryptographic calculation is represented by a benchmark score from the short-term specific tests according to the company's requirement. The higher the score, the faster the speed. Table 3.4 illustrates the fuzzy values of the three objective attributes for the four alternative clouds. And according to Equations (3.2) & (3.3), all these fuzzy values are converted into fuzzy ratings. Table 3.5 illustrates the fuzzy ratings corresponding to

Alter. Clouds \ Obj. Attr.	A_5 (%)	A_6 (ms)	A_7 ($scores$)
C_1	98.8, 99.2, 99.2, 99.8	50, 56, 56, 60	136.4, 136.4, 136.4, 136.4
C_2	99.2, 99.8, 99.8, 100	75, 83, 83, 90	102.5, 102.5, 102.5, 102.5
C_3	97.9, 98.5, 98.5, 99.1	180, 200, 200, 220	62.6, 62.6, 62.6, 62.6
C_4	98.2, 98.8, 98.8, 99.4	160, 173, 173, 190	69.8, 69.8, 69.8, 69.8

Table 3.4: The fuzzy values of the objective attributes for each alternative cloud

Alter. Clouds \ Obj. Attr.	A_5 ($\widetilde{r}_5$)	A_6 ($\widetilde{r}_6$)	A_7 ($\widetilde{r}_7$)
C_1	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
C_2	99.2, 99.8, 99.8, 100	55.56, 60.24, 60.24, 66.67	75.15, 75.15, 75.15, 75.15
C_3	97.9, 98.5, 98.5, 99.1	22.73, 25, 25, 27.78	45.89 45.89 45.89 45.89
C_4	98.2, 98.8, 98.8, 99.4	26.32, 28.90, 28.90, 31.25	51.17, 51.17, 51.17, 51.17

Table 3.5: The fuzzy ratings of the objective attributes for each alternative cloud

Table 3.4.

Step 3: Combining the fuzzy ratings for all the subjective attributes and the objective attributes, the decision makers are listed for an alternative cloud. For the sake of simplicity, we only list the decision makers for C_1 in Table 3.6. The Euclidean distance between the corresponding associated attribute pairs are computed for each decision maker with Equation (3.4). For DM_4 and DM_8, we can see that the ratings of the subjective associated attributes A_3 and A_4 are much lower than the ratings of the corresponding objective associated attributes A_5 and A_6. By Equation (3.4), the Euclidean distances of DM_4 and DM_8 are 120.02 and 113.11 respectively, which exceed the threshold of 107.48 (i.e., 80% of the maximum distance of 134.35). Thus, DM_4 and DM_8 are considered unreasonable and filtered out of the decision-maker list.

Decision makers \ Attributes	A_1	A_2	A_3	A_4	A_5	A_6	A_7
1	B.P&F	VP	B.G&VG	B.G&VG	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
2	B.P&F	G	P	F	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
3	G	VG	B.F&G	G	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
4	B.G&VG	B.F&G	B.VP&P	VP	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
5	G	G	B.F&G	B.P&F	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
6	P	B.VP&P	G	G	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
7	B.VP&P	F	B.F&G	B.G&VG	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
8	F	VG	B.VP&P	B.VP&P	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
9	B.F&G	B.F&G	B.G&VG	B.P&F	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100
10	B.P&F	B.F&G	B.F&G	B.P&F	98.8, 99.2, 99.2, 99.8	83.33, 89.29, 89.29, 100	100, 100, 100, 100

Table 3.6: The decision makers of the alternative cloud C_1

Step 4: According to the potential cloud user (i.e., the company)'s requirement, the importance weight for each attribute is given in the form of linguistic variables. According to Table 3.3, all the weights are mapped into the fuzzy weights which are then converted into the crisp weights by Equation (3.5). All the weights are illustrated in Table 3.7.

Step 5: According to Table 3.6 and Table 3.7, the fuzzy scores of C_1 based on the remaining eight decision makers are computed by Equation (3.6). Finally, by Equation (3.7), the final score of C_1 is 74.3.

Results: Table 3.8 illustrates the scores of the four alternative clouds. Row 2 presents the scores based on our model considering both the subjective attributes and the objective attributes. Row 3 and Row 4 respectively present the scores of considering only the subjective attributes and only the objective attributes according to the corresponding importance weights. The ranking computed based on our model is C_1, C_2, C_3, C_4. The ranking computed based on the subjective assessment only is C_2, C_1, C_3, C_4. And the ranking computed based on the objective assessment only is C_1, C_2, C_4, C_3.

From Table 3.8, we can see that, although the subjective assessment from cloud users places C_1 as the second best, C_1 is the best cloud service according to our model since our model considers both its subjective assessment and objective assessment. It should be noted that the gap in the scores between C_1 and C_2 based on our model is smaller than the gap of the scores between them based on considering only their objective assessments. That is because our model considers subjective assessment for some important aspects of a cloud service, which can hardly be measured by quantitative testing. Thus, in this case, C_1 is the best cloud service by comprehensively considering all the concerned performance aspects of a cloud service according to the potential cloud user (i.e., the company)'s requirements.

Attributes \ Weights	Linguistic Variable	Fuzzy Weights	Crisp Weights
A_1	Very High	7,10,10,10	0.2022
A_2	High	5,7,7,10	0.1585
A_3	Low	0,3,3,5	0.0601
A_4	Medium	2,5,5,8	0.1093
A_5	Medium	2,5,5,8	0.1093
A_6	High	5,7,7,10	0.1585
A_7	Very High	7,10,10,10	0.2022

Table 3.7: The weights of each attribute

Approaches	C_1	C_2	C_3	C_4
Sub.&Obj.	**74.3**	67.8	51.5	50.6
Sub.	54.4	**60.5**	51.8	47.0
Obj.	**96.6**	76.0	51.1	54.7

Table 3.8: The scores of each alternative cloud

3.4 Conclusion

In this chapter, we have proposed a novel model of cloud service selection by aggregating subjective assessments from cloud consumers and objective performance assessment from a trusted third party. We apply a fuzzy simple additive weighting system to normalize and aggregate all different types of subjective attributes and objective attributes of a cloud service, so that some specific performance aspects of a cloud service can also be taken into account according to potential cloud users' requirements. Furthermore, our model can identify and filter unreasonable subjective assessments. This makes the results based on our model more accurate and effective with less noise. Based on the analysis through a case study, our proposed model of cloud service selection has the following advantages:

(1) By considering subjective assessments from cloud consumers, our model takes into account some vital but qualitative performance aspects in the selection process of a cloud service as well as quantitative performance aspects.

(2) Our model considers the situation in the real world, where cloud users' subjective assessments are fuzzy in linguistic form as well as the importance weight for

each performance attribute. Thus, our model can effectively deal with the uncertainty of human languages in cloud service selection.

(3) According to the different concerns of potential cloud users for different subjective attributes and objective attributes, our model presents an overall performance score for a cloud service by aggregating all subjective assessments and objective assessments with less noise from unreasonable subjective assessments.

Chapter 4

Cloud Service Selection based on Contextual Assessments

Due to the diversity and dynamics of cloud services, selecting the most suitable cloud service has become a major issue for potential cloud consumers. Prior to cloud service selection, an evaluation of cloud services should be applied first. In the literature, there are two types of approaches which can be used to conduct such an evaluation. The first type of approaches is based on objective performance assessments from ordinary QoS (Quality-of-Service) value (e.g., service response time, availability and throughput) monitoring and predesigned benchmark testing. The second type of approaches is based on user subjective assessments which are usually extracted from user ratings for each concerned aspect of cloud services. In this type of approaches, cloud services are usually treated like traditional web services, thus some rating-based reputation systems [148, 87, 101] can be utilized for cloud service selection.

Nevertheless, these two types of cloud service evaluation approaches have their own limitations. That is because, firstly, objective performance assessment can only be carried out for the performance aspects which can be easily quantified. Conversely, objective assessment is not appropriate for those aspects which are quite hard to quantify, e.g., data privacy. On the other hand, subjective assessment has the risk of inaccuracy since users' subjective feelings are very likely to contain bias and not reflect the real situations of cloud performance. In addition, as cloud users who give subjective assessments are usually spread throughout the world, for any cloud service, the subjective feelings of a cloud user in a context (e.g., morning in Sydney) may be much different from those of another user in a different context (e.g., afternoon in Paris). Furthermore, there may be malicious users who give unreasonable subjective assessments to deceive others and/or benefit themselves in some cases. As a result, the accuracy of overall subjective assessment for cloud services can be significantly affected. Hence, a cloud service selection model, which can be used to not only aggregate different performance aspects of cloud services according to cloud consumers' various needs but also filter unreasonable user subjective assessments, is highly desirable.

To overcome the aforementioned drawbacks, this chapter proposes a novel context-aware cloud service selection model based on the comparison and aggregation of subjective assessments extracted from cloud user feedback and objective assessments from quantitative performance testing. The proposed model is extended and modified from our prior work presented in Chapter 3. In this new model, according to a potential cloud consumer's requirements, an objective assessment provided by some professional testing party is first applied as a benchmark to filter out biased or unreasonable subjective assessments. In order to guarantee the accuracy of such filtering, our work considers two assessment features (i.e., *location* and *time*) in contexts, which can commonly affect both objective assessments and subjective assessments. The process of filtering is based on the context similarity between objective assessments and subjective assessments, i.e., the more similar the context, the more reliable subjective assessment, so that the benchmark level is dynamically adjusted according to the corresponding context similarity, which is computed through a novel approach inspired by the SimRank Algorithm [69]. After such filtering, the final aggregated results can reflect the overall performance of cloud services according to potential users' personalized requirements and context.

In Section 4.1, we first introduce the contexts concerned in cloud service selection. After briefly introducing the preliminaries of our prior cloud selection model which does not consider assessment contexts, the details of our context-aware model are discussed in Section 4.2. Finally, Section 4.3 presents the experimental results to demonstrate the feasibility of our context-aware model.

4.1 Contexts in Cloud Service Selection

The definition of contexts usually varies in different application environments. In our cloud service selection model based on both objective assessments and subjective assessments, the context of an assessment for a cloud service refers to a group of values of the features of the assessment, which can affect the result of the assessment.

To give an example of the impact of a context, according to the objective statistics from CloudSleuth[1], the response time of a cloud service varies significantly under different worldwide QoS monitoring centers, and generally increases with the increasing distances between the cloud provider and these monitoring centers because of the increasing length of the network routes of cloud service delivery. Meanwhile, the monitoring results of response time can also be affected by the time of a day, in other words, how busy the cloud service and the network accessed by the monitoring centers for monitoring can vary at different times of a day. Therefore, both objective assessment and subjective assessment can be affected according to different assessment contexts. At the current stage of our work, we consider two assessment features (i.e., *location* and *time*) in our context-aware cloud service selection model.

In our prior cloud service selection model presented in Chapter 3, assessment contexts are not taken into account. However, in order to have a more accurate comparison between objective assessment and subjective assessment, the similarity between the contexts of objective assessments and subjective assessments should be considered. More similar contexts indicate the subjective assessments are given in the more similar situation with that of the given objective assessment, thus such subjective assessments are considered more reliable. Furthermore, in our prior model, a fixed threshold is used as the benchmark value for the objective assessment to filter out unreasonable subjective assessments. This threshold reflects how much the objective assessment is

[1] www.cloudsleuth.net

trusted. If the threshold is high, more subjective assessments are retained for the following aggregation process, which means more subjective assessments are considered reasonable, otherwise fewer subjective assessments are retained, which means fewer subjective assessments are considered reasonable. However, determining such a suitable fixed threshold is very difficult. Because the fixed threshold means the subjective assessments with different contexts are treated equally. If the threshold is determined too high, more noisy subjective assessments will be left in the final aggregated results. Conversely, if the threshold is too low, only a few subjective assessments are left so that only these few subjective assessments can affect the final aggregated results, and, as a consequence, the final aggregated results cannot reflect most users' subjective assessment. An intuitive solution to overcome this drawback is to adjust the threshold dynamically according to the context similarity between objective assessment and subjective assessment. The more similar the contexts, the more reliable subjective assessments, thus the threshold should be set higher for retaining more of such subjective assessments. On the contrary, if the contexts are less similar, then the threshold should be set lower to filter out more subjective assessments which are given in more different situations. Next, the details of computing such context similarity will be introduced.

4.1.1 Context Similarity

In [155], Tavakolifard *et al.* introduce a general idea of the calculation of context similarity based on the bipartite SimRank algorithm [69] for trust transferability among similar contexts in electronic transactions. In order to compute the similarity of assessment contexts in our context-aware cloud service selection model, we follow Tavakolifard *et al.*'s idea and propose a concrete approach for context similarity measurement. In details, our approach consists of two steps:

The ***first step*** is to compute the similarity between two values from the same assessment feature.

The ***second step*** is to model all contexts and their relevant assessment features as

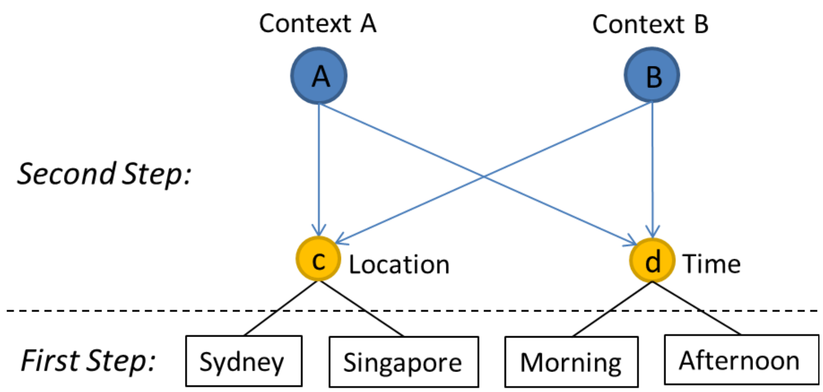


Figure 4.1: An Example of Two Contexts

a graph and compute the overall similarity between contexts.

Figure 4.1 illustrates an example of two contexts A (*Sydney, morning*) and B (*Singapore, afternoon*) belonging to two assessments. Each context contains two values for two assessment features (i.e., *location* and *time*). *Sydney* and *Singapore* are the values of the feature *location* for both contexts. Likewise, *morning* and *afternoon* are the values of the feature *time*.

In [155], Tavakolifard *et al.* only introduced how to compute overall context similarity (i.e., the second step) through the bipartite SimRank algorithm and did not present details on computing the similarity between two values from the same assessment feature (i.e., the first step). For each assessment feature, a specific comparator needs to be designed for computing similarity among the values of each feature. In our context-aware cloud service selection model, two features are considered in assessment contexts. Next, we first present a modified version of the bipartite SimRank algorithm according to our model, and then introduce the design of the comparators for *location* and *time*.

4.1.1.1 Modified Bipartite SimRank

The original bipartite SimRank algorithm [155] is modified to take different context comparators into account in our model. Let A and B denote two contexts and, $s(A, B)$

denote the similarity between A and B. If $A = B$, then $s(A, B) \in [0, 1]$ is defined to be 1. Let c and d denote assessment features for contexts A and B, and $s(c, d) \in [0, 1]$ denote the similarity between features c and d. Let $V_c(A)$ and $V_c(B)$ denote the values of the feature c in the contexts A and B respectively. Likewise, $V_d(A)$ and $V_d(B)$ denote the values of the feature d in the contexts A and B respectively. If $c = d$, then $s(c, d) = Cmp_c(V_c(A), V_c(B)) = Cmp_d(V_d(A), V_d(B)) \in [0, 1]$, where Cmp_c and Cmp_d are the comparators for the features c and d.

Now, A, B and c, d can be formed to a directed graph pointing from contexts to features. If we take Figure 4.1 as an example, we have that $A = (Sydney, morning)$, $B = (Singapore, afternoon)$, $c = location$, $d = time$, $V_c(A) = Sydney$, $V_c(B) = Singapore$, $V_d(A) = morning$ and $V_d(B) = afternoon$. In the directed graph, $I(v)$ and $O(v)$ denote the set of in-neighbors and out-neighbors of v respectively, where v is a node in the graph. $I_i(v)$ denotes an individual in-neighbor of v for $1 \leq i \leq |I(v)|$, and $O_i(v)$ denotes an individual out-neighbor of v for $1 \leq i \leq |O(v)|$.

Now we have the recursive equations: for $A \neq B$,

$$s(A, B) = \frac{C}{|O(A)||O(B)|} \sum_{i=1}^{|O(A)|} \sum_{j=1}^{|O(B)|} s(O_i(A), O_j(B)), \tag{4.1}$$

and for $c \neq d$,

$$s(c, d) = \frac{C}{|I(c)||I(d)|} \sum_{i=1}^{|I(c)|} \sum_{j=1}^{|I(d)|} s(I_i(c), I_j(d)), \tag{4.2}$$

where $C \in (0, 1)$ is a constant which can be considered as either a confidence level or a decay factor. In the full version [68] of Jeh *et al.*'s paper [69] proposing bipartite SimRank, they argue that the constant C can be viewed as the bases of exponential functions whose only purpose is to map distances to finite intervals. Although the values of similarity can be affected by C, the relative results of similarity is still retained. Hence, for the sake of efficiency, we follow Jeh *et al.*'s setting to set $C = 0.8$ in our model. In addition, Jeh *et al.* have proven that a simultaneous solution $s(*, *) \in [0, 1]$

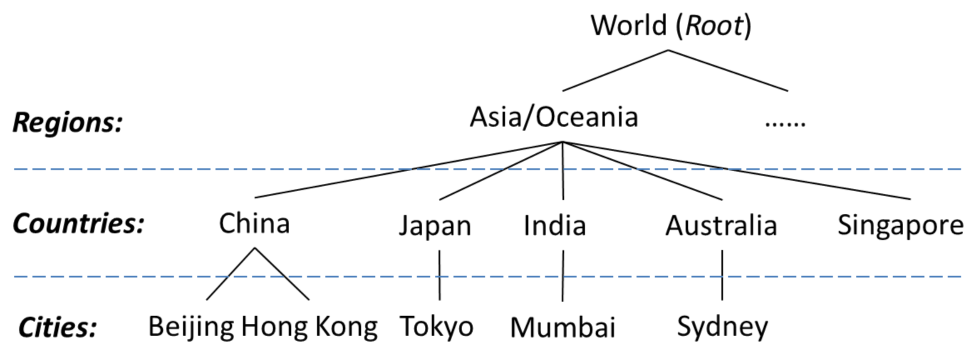


Figure 4.2: A Geographical Hierarchy

to the recursive equations (4.1) and (4.2) always exists and is unique.

4.1.1.2 Design of Comparators

According to each assessment feature, a corresponding comparator needs to be designed and applied in the above modified bipartite SimRank algorithm. In our model, two assessment features are considered, i.e., *location* and *time*.

Similarity of Locations:

The effect for both objective assessment and subjective assessment of cloud services is usually caused by the delay of the Internet communication between the locations of where the assessments are given and the target cloud service. In order to precisely model such an effect, the Internet topology between these parties should be first determined. However, such a topology should be created by some domain experts, and is out of the scope of our work. For the sake of simplicity, in this chapter, we use geographical locations instead of the Internet locations. That is because the distance between two nodes in the Internet is commonly determined by their geographical locations. We introduce a similarity measurement based on a hierarchical ontology structure [187] for the assessment feature *location* in our model.

According to the real monitoring data from CloudSleuth, we establish a geographical hierarchy according to the order of $regions \rightarrow countries \rightarrow cities$. Figure 4.2

illustrates the Aisa/Oceania part of the hierarchy. In order to measure the similarity between any two nodes in the hierarchy, we apply Zhang *et al.*'s hierarchy-based approach of similarity measurement [187]. Let D denote the depth of the deepest common ancestor of two nodes n and n'. For example, the deepest common ancestor of *Beijing* and *Tokyo* is *Asia/Oceania* in Figure 4.2, thus $D(Beijing, Tokyo) = 1$. The smaller D represents the deepest common ancestor of the two nodes is on the upper layer of the hierarchy, which means the two nodes are fallen into a more general classification, thus are less similar. Conversely, a larger D means the two nodes are fallen into a more concrete classification, thus are more similar. Hence, a monotonically increasing hyperbolic tangent function [187] is defined to model this trend:

$$Cmp(n, n') = \frac{e^{\alpha D(n,n')} - e^{-\alpha D(n,n')}}{e^{\alpha D(n,n')} + e^{-\alpha D(n,n')}}, \tag{4.3}$$

where $Cmp(n, n')$ represents the similarity comparator returning the similarity value between n and n'; $\alpha \in (0, 1)$ is a constant. Here, we follow Zhang *et al.*'s setting to set $\alpha = 0.4$, which is considered optimal according to their experimental results.

Similarity of Time:

In practice, the reasons why the different times of a day can affect both objective assessment and subjective assessment of cloud services are quite diverse and complicated, where the main reason for such an effect is how busy the networks used by users to access cloud services are. However, the extent of how busy networks are varies frequently according to different users' situations, thus it is also quite hard to quantitatively measure such changes.

Hence, in our model of context-aware cloud service selection, we divide 24 hours of a day into two time intervals. When a potential cloud user asks for cloud service selection, he/she needs to specify in what period of time he/she hopes to frequently employ the selected cloud service. The assessments given within that period of time are considered more reliable for the potential user, and the assessments given within

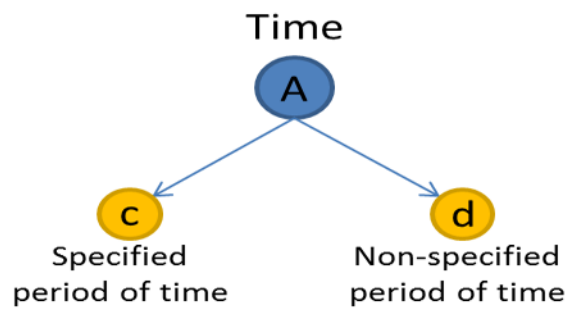


Figure 4.3: Similarity of Time

the non-specified period of time are considered less reliable for the user. Therefore, in our model every subjective assessment contains a time stamp to identify the time when the assessment is given. We assume that such assessments are required to represent users' subjective judgement at that time only. To this end, we propose an incentive mechanism in Chapter 7 for giving cloud users incentives to provide subjective assessments regularly. Due to the incentive mechanism, most cloud users will give such subjective assessments with time stamps.

In our model, the assessment feature *time* has two states, i.e., specified and non-specified. The similarity between these two states can be computed through the basic bipartite SimRank algorithm. Figure 4.3 illustrates the graph of similarity of the two states. Then, the similarity between *specified period of time* and *non-specified period of time* can be computed through Equations (4.1) and (4.2).

It should be noted that, except *location* and *time*, there are some other assessment features which can also affect the assessment results for some reasons (e.g., the Internet service providers). The similarity among the values of such a feature should be computed through a specific designed comparator. And the modified bipartite SimRank algorithm introduced above can be applied with any further comparator.

4.2 Context-aware Cloud Service Selection

In our prior cloud service selection model introduced in Chapter 3, a framework is proposed for cloud service selection based on both cloud user feedback and objective performance benchmark testing. This framework is composed of four components, namely, (1) *cloud selection service*, (2) *benchmark testing service*, (3) *user feedback management service*, and (4) *assessment aggregation service*, where *cloud selection service* is in the higher layer of the framework to command the others in the lower layer. Through defining the associated performance attributes (refer to Chapter 3), our proposed cloud service selection approach consist of five steps: *1) converting the values of subjective attributes into ratings, 2) converting the values of objective attributes into ratings, 3) filtering out unreasonable subjective assessments, 4) computing the importance weight for each attribute*, and *5) aggregating all attributes*. In our context-aware model, we basically follow the framework and selection procedure introduced in Chapter 3, and made some modifications due to considering assessment contexts.

In our context-aware cloud service selection model, we assume there are plenty of *benchmark testing agents* spread around the world providing *benchmark testing services*. When a potential cloud user asks for selecting the most suitable cloud service, according to his/her situation, he/she needs to specify which agents should be selected to offer objective assessments for all alternative cloud services. Then, the cloud service selection will be processed independently according to each agent. For each *benchmark testing agent*, the *cloud selection service* asks the *user feedback management service* to provide the subjective assessments for all alternative cloud services from the cloud users all over the world. Then, all the subjective assessments are classified according to their contexts (i.e., *location* and *time*). As the nodes in the deepest level of our geographical hierarchy are cities, the location of each subjective assessment is set as the nearest city to the real location specified in the assessment in the hierarchy. And due to time differences among cloud users all over the world, the time specified in every assessment is converted into one standard time.

Assume that there are l locations shown in all the subjective assessments. As there are only two states for the assessment feature *time* in our model, i.e., *specified period of time* and *non-specified period of time*, all the subjective assessments are classified into $2l$ groups. Then, according to the potential user's requirement, the *benchmark testing agent* provides an objective assessment with contextual information (e.g., objective performance assessment in the *morning* of *Sydney*).

The process of the comparison and aggregation of the objective assessment and the subjective assessments is the same as that of our prior work without the consideration of assessment contexts, except the importance weight setting and changing a fixed threshold to a group of dynamical thresholds. Such thresholds are computed as follows:

Step 1: The potential user first sets the importance weights on how much to trust objective assessment or subjective assessment through linguistic variables. Then, through a mapping [26], linguistic weights are converted into fuzzy weights, which are denoted as $\widetilde{W_o}$ and $\widetilde{W_s}$ for objective assessment and subjective assessment respectively. Then, the potential user sets the importance weight for each objective or subjective attribute, denoted as $\widetilde{W_i}$, where $i = 1, \cdots, s+o$. After that, W_i is the normalized weight of each attribute, which is computed as follow:

$$
\begin{aligned}
W_i &= \frac{d(\widetilde{W_s})}{d(\widetilde{W_s}) + d(\widetilde{W_o})} \times \frac{d(\widetilde{W_i})}{\sum_{i=1}^{s} d(\widetilde{W_i})}, \; i = 1, \cdots, s, \\
W_i &= \frac{d(\widetilde{W_o})}{d(\widetilde{W_s}) + d(\widetilde{W_o})} \times \frac{d(\widetilde{W_i})}{\sum_{i=s+1}^{s+o} d(\widetilde{W_i})}, \; i = s+1, \cdots, s+o.
\end{aligned} \tag{4.4}
$$

Step 2: Let g_o denote the context of the objective assessment, and g_v denote the context of each group of the subjective assessments in the total $2l$ groups, where $1 \leq v \leq 2l$. Through the approach introduced in Section 4.1, the similarity between each g_v and g_o is computed and denoted as $s_v(g_v, g_o)$.

Step 3: In order to offset the effect caused by the constant C in the modified bipartite SimRank algorithm, let $s_o(g_o, g_o)$ denote the similarity between the contexts of the objective assessment and itself, and E_{dis} denote the theoretical maximum Euclidean distance between corresponding objective associated attributes and subjective associated attributes according to our model. The filtering threshold R_v for the subjective assessment group with the context g_v is weighted by the attribute importance and the context similarity as follow:

$$R_v = (1 - \frac{d(\widetilde{W_o})}{d(\widetilde{W_s}) + d(\widetilde{W_o})}) \times \frac{s_v(g_v, g_o)}{s_o(g_o, g_o)} \times E_{dis}, \tag{4.5}$$

where $v = 1, \cdots, 2l$. From the above equation, we can see when the potential user trusts objective assessment more, R_v will become smaller, so that more subjective assessments are considered unreasonable and will be filtered out. In addition, when the context similarity $s_v(g_v, g_o)$ becomes lower, R_v will become smaller. That means the subjective assessments are given in a more different situation with that of the objective assessment, thus such subjective assessments are considered less reliable and will be filtered out more rigorously. Finally, the rest of the subjective assessments after such filtering and the objective assessment are aggregated to reflect the overall performance of a cloud service more accurately.

4.3 Experiments

4.3.1 Experiment Setup

In our experiments, there are three subjective attributes, i.e., *cloud provider reputation on privacy (A_1), after-sales services (A_2)*, *service response time (A_3)*, and two objective attributes, i.e., *service response time (A_4)* and *CPU performance (A_5)*, where *service response time* A_3 and A_4 are the associated attribute pair.

In order to evaluate our context-aware cloud service selection model, two kinds of data are required, i.e., subjective ratings from cloud users, and objective results of QoS monitoring and benchmark testing. In our experiments, we collect the data of *response* *time* A_4 from CloudSleuth and the data of benchmark scores of *CPU performance* A_5 from CloudHarmony[2] for 59 real cloud services. To the best of our knowledge, there is no data set of cloud user ratings published for these 59 cloud services. Hence, we simulate user ratings of the attributes A_1, A_2 and A_3 according to the collected objective data (i.e., A_4 and A_5). In details, the ratings of A_1 and A_2 are randomly generated, and the normal ratings of A_3 are generated according to the ranking of the real data of *response time* in A_4. Then, some biased ratings are added into the normal ratings of A_3 to simulate the ratings from the users who are in different contexts with that of objective assessments. Here, a bias level denoted as BL is set to represent how much the biased ratings deviate from the normal synthetic ratings of A_3, where $BL = 1, \cdots, 8$ since a rating scale of 1-9 is employed in our model. Moreover, a biased rating percentage denoted as BRP is set to represent how many biased ratings there are in all the subjective ratings.

We assume that all the subjective ratings are from the cloud users belonging to two different contexts. The one context is (*Sydney, afternoon*) which is also the context of the objective assessment in our experiments, and the other context is (*Hong Kong, morning*). According to the algorithm introduced in Section 4.1, the similarity of the two contexts is 0.4714. Thus, two thresholds are computed for the comparison of subjective assessment and objective assessment according to different importance weights (i.e., $\widetilde{W}_o$ and $\widetilde{W}_s$).

4.3.2 Evaluation Metric

In our experiments, we first generate 1000 normal ratings for the attributes A_1, A_2 and A_3 through the way introduced above, and then replace some proportion of normal ratings with biased ratings. Here, the original normal rating matrix is denoted as M_o, and the corresponding processed rating matrix including biased ratings is denoted as M_b.

[2]www.cloudharmony.com

As M_o is generated according to the objective assessment, the ratings in M_o are considered to be very accurate. Thus, the final aggregated result for each alternative cloud service without filtering between subjective assessment and objective assessment is considered very accurate in representing the overall performance of each cloud service. Here, $R(M_o)$ denotes the ranking of all the 59 cloud services based on such aggregated results without filtering according to M_o. $R_f(M_b)$ denotes the ranking of the cloud services based on our prior cloud service selection model without contextual assessments according to M_b without the consideration of assessment contexts; $R_c(M_b)$ denotes the ranking of the cloud services based on our context-aware cloud service selection model according to M_b with dynamic threshold filtering. $R_{sim}(*, *)$ denotes the similarity between two ranking lists. If $R_{sim}(R(M_o), R_c(M_b)) > R_{sim}(R(M_o), R_f(M_b))$, that means our context-aware model is more effective than our prior model.

In our experiments, $R_{sim}(*, *)$ is calculated through the *Kendall tau rank distance* [36] which is a common metric to measure the distance between two rankings through counting the number of pairwise disagreements between the two rankings. Here, we use the function $corr()$ provided in $Matlab$ to compute the normalized *Kendall tau distance* which lies in the interval $[-1, 1]$, where 1 means two rankings are in the same order, and -1 means two rankings are in the opposite order.

Importance Weights	BRP	Ranking Similarity \ BL	4	5	6	7	8
$W_s = High, W_o = High$	20%	$R_{sim}(R(M_o), R_f(M_b)) \times 100$	84.5235	88.9254	92.9106	92.9663	92.9837
$W_1 = VeryHigh$		$R_{sim}(R(M_o), R_c(M_b)) \times 100$	**90.9944**	**93.3161**	**94.8105**	**94.8335**	**94.9709**
$W_2 = High$	50%	$R_{sim}(R(M_o), R_f(M_b)) \times 100$	84.4224	88.9060	92.7962	92.6966	92.8795
$W_3 = Medium$		$R_{sim}(R(M_o), R_c(M_b)) \times 100$	**93.7033**	**93.6175**	**96.4285**	**96.3141**	**96.4009**
$W_4 = VeryHigh$	70%	$R_{sim}(R(M_o), R_f(M_b)) \times 100$	84.2947	89.0061	92.9147	92.8943	92.7497
$W_5 = High$		$R_{sim}(R(M_o), R_c(M_b)) \times 100$	**93.2850**	**93.8473**	**95.4045**	**95.3478**	**95.4597**
$W_s = High, W_o = VeryHigh$	20%	$R_{sim}(R(M_o), R_f(M_b)) \times 100$	92.4678	93.1859	93.9019	93.9888	93.9055
$W_1 = Medium$		$R_{sim}(R(M_o), R_c(M_b)) \times 100$	**94.3121**	**95.1874**	**95.2651**	**95.3754**	**95.3366**
$W_2 = High$	50%	$R_{sim}(R(M_o), R_f(M_b)) \times 100$	92.3825	94.1011	93.8963	93.9423	93.8958
$W_3 = High$		$R_{sim}(R(M_o), R_c(M_b)) \times 100$	**94.7589**	**96.3560**	**95.9770**	**96.0603**	**96.0587**
$W_4 = VeryHigh$	70%	$R_{sim}(R(M_o), R_f(M_b)) \times 100$	91.4525	93.1057	92.9219	92.9775	92.9704
$W_5 = Medium$		$R_{sim}(R(M_o), R_c(M_b)) \times 100$	**93.4980**	**95.3238**	**95.0511**	**95.1287**	**95.2288**

Table 4.1: Accuracy Comparison based on Ranking Similarity

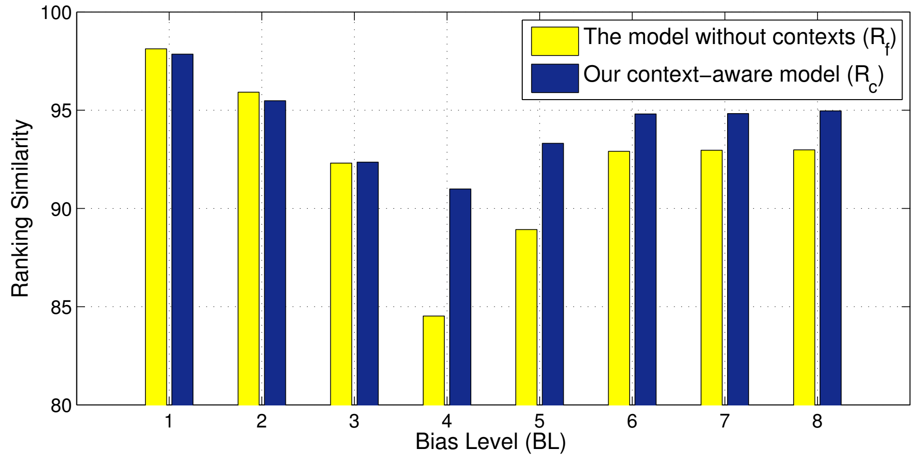


Figure 4.4: Ranking Similarity when BRP = 20%

4.3.3 Experimental Results

In our experiments, the importance weight for each attribute is randomly selected. According to our experiments, the importance weights do not affect the trend of our experimental results, that is, our context-aware cloud service selection model is more effective. Table 4.1 shows part of the experimental results for the 59 real cloud services based on two settings of importance weights. A larger value indicates better ranking accuracy. In order to more accurately simulate the ratings from real cloud users in our experiments, every value in Table 4.1 is the average ranking similarity computed based on every 100 different groups of M_o and M_b. And each group of data is generated independently. Thus, the generality of experimental data can be kept in our experiments. Table 4.1 shows that, among different experimental conditions (i.e., different BLs and BRPs), our context-aware cloud service selection model performs better than our prior model without the consideration of contexts. And our context-aware model can achieve approximately 1.5% to 9% improvements.

Table 4.1 shows that our context-aware cloud service selection model based on dynamic threshold filtering can more effectively deal with the effect of biased subjective ratings than our prior cloud service selection model in different conditions (i.e., different BLs and BRPs) except the conditions that $BL = 1, 2$ or 3. That is because, in the real world, cloud users' subjective assessment for a cloud service cannot perfectly match the objective assessment of the cloud service due to users' different preferences. However, users' subjective assessment should not be far off from objective assessment. For this consideration, in our experiments, every individual synthetic normal subjective rating does not perfectly match the objective assessment, and may have a random small deviation (up to 3). If the deviation (i.e., BL) between biased ratings and normal ratings is too small, such biased ratings are very likely to be considered as normal ratings since such a small deviation should not be detected as the deviation between biased ratings and normal ratings. That leads to the fact that our experimental results in the conditions of $BL = 1, 2$ or 3 may be opposite since such biased ratings with small deviations cannot be detected in our experimental setting. However, in the other conditions of any BRP and $BL = 4, \cdots, 8$, the trend of our experimental results is the same. Figure 4.4 illustrates such an example when $BRP = 20\%$.

4.4 Conclusion

This chapter has proposed a novel model of context-aware cloud service selection based on comparison and aggregation of subjective assessments from cloud consumers and objective assessments from quantitative QoS monitoring and benchmark testing, which is extended from our prior work.

The new model considers contextual subjective or objective assessments in cloud service selection, and uses objective assessment as a benchmark to filter out unreasonable subjective assessments. The process of such filtering is based on a group of dynamic thresholds which are determined by the similarity between the contexts of subjective assessments and objective assessments.

In our model, we have considered two assessment features *location* and *time*, both of which would greatly affect the results of cloud service evaluation and selection. A novel approach is proposed to compute the context similarity based on these two features. Our similarity computation approach can be easily extended for any other type of assessment features.

Our experimental results have demonstrated that our context-aware model performs better than our prior cloud selection model which has no consideration of assessment contexts. Hence, the final aggregated results of cloud services based on our context-aware model can more accurately reflect the overall performance of cloud services according to cloud consumers' requirements and contexts.

Chapter 5

Evaluating Cloud Users' Credibility of Providing Assessments

As introduced in Chapters 3 and 4, cloud service selection is usually based on subjective assessments (e.g., subjective ratings) from ordinary cloud consumers and/or objective assessments (e.g., benchmark testing results) from professional cloud testing parties. Whichever type of approaches is adopted, the credibility of cloud users providing assessments has a strong influence on the effectiveness of cloud service selection.

In cloud environments, cloud users can be generally classified into two classes according to the different purposes of consuming cloud services. The first class comprises ordinary cloud consumers whose purpose is to consume a cloud service having high quality performance and spend as little money as possible. They usually offer subjective assessment of cloud services through user feedback. The second class comprises professional cloud performance monitoring and testing parties whose purpose is to offer objective assessment of cloud services to potential cloud consumers for helping them select the most suitable cloud services. In general, objective assessment is considered more reliable than subjective assessment due to scientific and statistical analysis. However, objective assessment cannot be fully trusted since the parties who offer assessments may provide untruthful assessments due to their interest. And because there are only a small number of organizations which carry out objective testing of cloud services at present, cloud service selection based on the aggregation of both subjective assessment and objective assessment should be more effective than either type of approaches alone. To the best of our knowledge, there are no prior approaches in the literature, which can evaluate the credibility of both types of assessments in cloud environments.

In this chapter, we propose a novel model for evaluating cloud users' credibility of providing subjective assessments or objective assessments, where subjective assessments are from ordinary cloud consumers (called Ordinary Consumers, *OC* for short), and objective assessments are from professional cloud performance monitoring and testing parties (called Testing Parties, *TP* for short). Our model is based on two classes of cloud users (i.e., *OC*s and *TP*s). The credibility of OCs and TPs providing subjective assessments or objective assessments is respectively represented by *trustworthiness* of OCs and *reputations* of TPs.

For an *OC*, an authority center computes the *relative trustworthiness* of the other *OC*s who consume the same cloud services as the OC. Relative trustworthiness represents other OCs' trustworthiness from the OC's prospect. The relative trustworthiness can also be affected by the difference of variation trend between the other *OC*'s subjective assessments and *TP*s' objective assessments over time. Then, the authority center selects the OCs who are considered trustworthy enough by the OC as his/her virtual neighbors according to all the relative trustworthiness values. The neighborhood relationships of all the OCs form a social network. The global trustworthiness of an OC on how truthful he/she provides subjective assessment is computed based on the number of OCs who select him/her as their virtual neighbor.

In the meantime, the reputation of a *TP* on providing truthful objective assessment is modeled in a different way based on the difference among the *TP*'s objective assessments, the majority of objective assessments from other TPs and the majority of subjective assessments from *OC*s. That implies that the trustworthiness of OCs and the reputations of TPs can be influenced by each other. For this reason, our model can resist collusion among cloud users providing untruthful assessments to some extent. Through our model, a successful collusion attack would become very difficult in practice since a large number of cloud users would have to be involved in such collusion.

We conduct a series of experiments to evaluate the performance of our model under different circumstances, where some cloud users (OCs and/or TPs) behave maliciously according to different strategies. In contrast to the existing user credibility evaluation model which is based on subjective ratings only, our experimental results show that our model can significantly improve the accuracy of evaluating user credibility, and enhance the resistance capability of user collusion in cloud environments, because the evaluation of cloud users' credibility in our model is based on both subjective assessments and objective assessments.

In a nutshell, the motivation of proposing such a credibility evaluation model is to enhance the effectiveness of cloud service selection approaches. If cloud users are considered more credible, their assessments should be considered more valuable in cloud service selection. Section 5.1 presents the details of our credibility evaluation model. Section 5.2 presents the experimental results which validate the feasibility of our proposed model.

5.1 The Credibility Evaluation Model

In this section, we first introduce the framework of our proposed model for evaluating cloud users' credibility, and then present the details of our model.

5.1.1 The Framework for User Credibility Evaluation

Fig. 5.1 illustrates the framework of our model consisting of two sub models, each of which targets one class of cloud users, i.e., OCs or TPs. In our framework, subjective assessments for cloud services are extracted from ratings submitted by ordinary consumers, and objective assessments are offered by testing parties using their own benchmark testing tools. After that, subjective assessments and objective assessments will

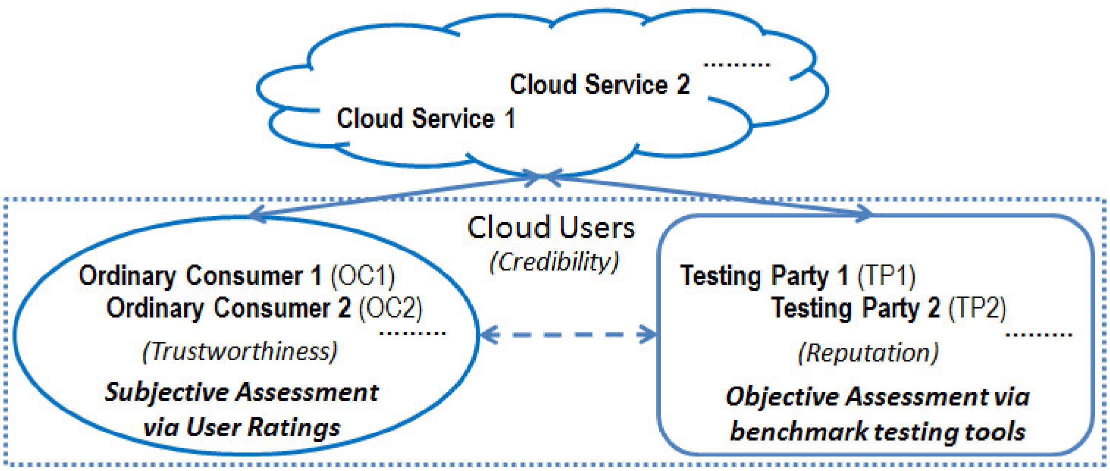


Figure 5.1: The Framework of the Credibility Evaluation Model

be aggregated in the further cloud service selection process, e.g., the process specified in [131]. In our framework, there is an authority center which is in charge of managing assessments of cloud services and evaluating the trustworthiness and reputation of every OC and TP on how truthfully they provide assessments.

The basic idea behind our work is that, an OC having higher trustworthiness means that his/her subjective assessments are more similar to those of many other OCs, and the variation trend of his/her subjective assessments over time are more similar to that of objective assessments from the TPs having high reputations. In addition, a TP with a higher reputation means that its objective assessments are more similar with the majority of objective assessments from other TPs and the majority of subjective assessments from OCs. It should be noted that, in practice, objective assessments are usually more reliable than subjective assessments since subjective assessments may contain users' subjective bias but objective assessments with scientific and statistical analysis do not. Thus, a TP should not be simply considered as an OC with more credibility since it is hard to quantitatively determine how much more credible a TP is compared to an OC. That is the reason why two sub models are applied to respectively compute OCs' trustworthiness and TPs' reputations.

Without loss of generality, we focus on the situation, where both subjective assessments and objective assessments evaluate one performance aspect of cloud services. For example, the *response time* of a cloud service can be quantitatively tested by TPs. Meanwhile, an OC consuming the same cloud service can also give his/her subjective ratings for the service response time by sensing how long the cloud responds to his/her requests. The situation of considering multiple performance aspects can be modeled based on evaluating user credibility on every performance attribute separately. We assume that all assessments are given in similar circumstances to avoid the situation, where two assessments given by two credible cloud users for a cloud service respectively, are quite different because the circumstances, under which the users give the assessments are quite different from each other (e.g., two cloud users live geographically far away).

5.1.2 The Sub Model for Computing Trustworthiness of OCs

The basic idea of evaluating trustworthiness of OCs in this sub model is that, an OC is considered trustworthy to provide truthful subjective assessments if there are many other OCs or TPs whose subjective assessments or objective assessments are similar to his/hers. To this end, we improve Zhang *et al.*'s approach [189]. Firstly, a series of multiple ratings commonly employed by most rating systems for cloud services, such as CloudReviews[1], are employed instead of binary ratings (i.e., "0" and "1" ratings) in Zhang *et al.*'s work to express OCs' subjective assessments. Secondly, in our model, the trustworthiness of an OC can also be influenced by the reputations of TPs. If the variation trend of an OC's subjective assessments over time is more similar to those of objective assessments from TPs having high reputations, the OC's subjective assessments are considered more trustworthy. For example, suppose that an OC gives a subjective assessment sequence $\{1, 4, 2\}$ for a cloud service over time through a normalized rating system with the scale of $1 - 5$. Then a TP gives its objective assessment sequence $\{3, 5, 4\}$ over the same time period. Although their assessments

[1]www.cloudreviews.com

are quite different, the variation trends of their assessments are very similar. That is probably because the OC may have subjective bias (e.g., giving lower or higher ratings according to their preference), but the similar variation trend indicates that such subjective assessments are very likely to be given according to the OC's real observation. Hence, our model considering variation trends of objective assessments from TPs as a 'semi-trusted judge' can more accurately evaluate trustworthiness of OCs, as supported in the literature [194]. Finally, in our model, we apply the PageRank algorithm [123] to compute *global trustworthiness* of OCs instead of Zhang *et al.*'s method. Through the analysis of a case introduced in Section 5.1.2.2.2, our method is fairer than Zhang *et al.*'s. The experimental results also support this conclusion. In the following sections, we first introduce how to adopt a multiple rating system instead of a binary rating system, and then present the details of the sub model.

5.1.2.1 Distance Measurement between Multiple Ratings:

In [189], a binary rating system is applied. In order to make our model more practical, a multiple rating system is applied instead. In our prior work [131], we proposed a cloud service selection model based on a fuzzy rating system in order to deal with the uncertainty of human subjective perception. Cloud users' subjective assessments are first expressed in linguistic ratings (e.g., "good" or "bad") which are close to human language descriptions. Then a mapping from linguistic ratings to fuzzy numbers is applied. Table 5.1 lists the mapping applied in [131], which is also frequently used in prior literature, such as [92, 26]. In Table 5.1, each linguistic rating is mapped to a trapezoidal fuzzy number, denoted as $\widetilde{A} = (a, b, c, d)$, where $a < b < c < d$ are real numbers. A defuzzification method is defined to convert fuzzy numbers into crisp numbers, i.e., the defuzzified value of $\widetilde{A}$ is its *signed distance*: $d(\widetilde{A}) = \frac{1}{4}(a+b+c+d)$.

In this sub model, we apply the rating system defined in Table 5.1 to express OCs' subjective assessments. In order to compare two ratings, we adopt the approach proposed by Li and Wang [88], which maps the *rating space* into a *trust space*, to measure the distance between two ratings. As shown in Table 5.1, fuzzy ratings are

Linguistic Ratings	**Fuzzy Ratings**	**Crisp Ratings**	**Normalized Ratings** (r_i)
Very low (VL)	$(0, 0, 0, 3)$	0.75	0
Low (L)	$(0, 3, 3, 5)$	2.75	0.235
Medium (M)	$(2, 5, 5, 8)$	5	0.5
High (H)	$(5, 7, 7, 10)$	7.25	0.765
Very High (VH)	$(7, 10, 10, 10)$	9.25	1

Table 5.1: A Multiple Fuzzy Rating System

first converted into crisp ratings through the signed distance defuzzification method. Then, the crisp ratings are normalized into the interval $[0, 1]$ according to their values. The interval $[0, 1]$ is partitioned into k mutually exclusive ratings, denoted as $r_1, r_2, \cdots, r_i, \cdots, r_k$, where k is the size of the rating space and $1 \leqslant i \leqslant k$. In our setting, $k = 5$ and the five *normalized ratings* in $[0, 1]$ are shown in Table 5.1. A trust space for a service is defined as a triple $T = \{(t, d, u) | t \geqslant 0, d \geqslant 0, u \geqslant 0, t + d + u = 1\}$. Through Bayesian Inference and the calculation of *certainty* and *expected probability* based on a number of sample ratings, normalized ratings can be put into three intervals, i.e., for a normalized rating $r_i \in [0, 1]$, we have

$$r_i \text{ is } \begin{cases} distrust, & \text{if } 0 \leqslant r_i \leqslant d; \\ uncertainty, & \text{if } d < r_i < d + u; \\ trust, & \text{if } d + u \leqslant r_i \leqslant 1. \end{cases}$$

A rating in the *distrust* range means the consumer who gave this rating deems that the service provider did not provide the service with committed quality, and we have a contrary conclusion when a rating is in the *trust* range. A rating in the *uncertainty* range means the consumer is not sure whether the service is provided with committed quality. Here, we call such a range a ***trust level***. In our model, the authority center can first employ some credible OCs and ask them to provide truthful subjective assessments for cloud services. After collecting sufficient subjective assessments for a period of time, the center learns such t, d and u from these samples. Note that, other multiple rating systems can also be applied in our model by determining suitable trust levels according to the corresponding rating samples.

5.1.2.2 The Trustworthiness of OCs:

The computation of the trustworthiness of an ordinary consumer OC_A is based on Zhang *et al.*'s approach [189] without its drawback discussed in Section 5.1.2.2.2. The detailed process consists of two steps: in Step 1, the authority center computes all the other OCs' *relative trustworthiness* based on OC_A's own experience, and selects a fixed number of top OCs according to the descending order of all their relative trustworthiness values, where these top OCs are considered as OC_A's virtual neighbors. Here, relative trustworthiness represents other OCs' trustworthiness from OC_A's prospect. In Step 2, all these neighborhood relationships form a virtual social network, based on which, the *global trustworthiness* of all OCs are computed.

The details of these two steps are introduced below:

5.1.2.2.1 Step 1. Computing Relative Trustworthiness of *OC*s: Suppose there are two ordinary consumers denoted as OC and OC', both of whom consume a group of cloud services, denoted as $\{s_1, s_2, \cdots, s_i, \cdots, s_l\}$. The relative trustworthiness of OC' based on OC is denoted as $RTr(OC \sim OC')$, where $OC \neq OC'$, and is computed as follows:

$$\begin{aligned}&RTr(OC \sim OC') = \overline{R_{TP}(OC')} \times \\ &[\omega \times S_{pri}(OC \sim OC') + (1-\omega) \times S_{pub}(OC' \sim ALL)].\end{aligned} \tag{5.1}$$

The details in Eq. (5.1) are introduced below:

1. $S_{pri}(OC \sim OC')$ **(private similarity between *OC* and *OC'*)**: All ratings for a service s_i rated by OC and OC' are ordered into two rating sequences, denoted as $\overrightarrow{r_{OC,s_i}}$ and $\overrightarrow{r_{OC',s_i}}$ respectively, according to the time when the ratings are provided. The rating sequences are then partitioned in mutually exclusive time windows. The length of each time window may be fixed or determined by the frequency of the submitted ratings for s_i. Moreover, it should be considerably small so that the performance of s_i can hardly change in a time window. After that, a pair of ratings $(r_{OC,s_i}, r_{OC',s_i})$, each of which is from its own rating sequence, is said to be *correspondent* only if they are given in the same time window. If there are more than one correspondent rating pairs in a time window, the most recent r_{OC,s_i} and r_{OC',s_i} are put together as the correspondent rating pair for this time window. This setting can prevent a malicious cloud user from providing numerous untruthful assessments in a short period of time in order to manipulate cloud services' reputations.

Let N_{s_i} denote the total number of correspondent rating pairs for s_i in all the time windows, then the total number of such pairs for all cloud services is computed as $N_{all} = \sum_{i=1}^{l} N_{s_i}$. Recall the ***trust level***s introduced above. If the two ratings of a correspondent rating pair are in the same trust level, such a pair is said *positive*, otherwise *negative*. Thus, if there are N_p positive pairs, then the number of negative pairs is $N_{all} - N_p$. A positive correspondent rating pair means the ratings submitted by OC and OC' in this time window are similar; A negative pair means they are quite different. In Eq. (5.1), $S_{pri}(OC \sim OC')$ is called the *private similarity* of OC' which presents the similarity between the ratings provided by OC and OC', and computed as follows:

$$S_{pri}(OC \sim OC') = \frac{N_p}{N_{all}}. \tag{5.2}$$

2. $S_{pub}(OC' \sim ALL)$ **(public similarity between *OC'* and all other *OC*s)**: If there are insufficient correspondent rating pairs between OC and OC', OC''s *public similarity*, denoted as $S_{pub}(OC' \sim ALL)$ in Eq. (5.1), should be calculated. Here, we follow Zhang *et al.*'s idea, i.e., the public similarity of OC' depends on the similarity between his/her ratings and the majority of ratings submitted by the other OCs. In each time window, the most recent r_{OC',s_i} and the average of the other ratings submitted by the other OCs for s_i are put together as a correspondent rating pair, denoted as $(\overline{r_{s_i}}, r_{OC',s_i})$. Suppose the total number of such correspondent rating pairs for all cloud services is N'_{all}, where there are N'_p positive pairs. The public similarity of OC' is computed as follows:

$$S_{pub}(OC' \sim ALL) = \frac{N'_p}{N'_{all}}. \tag{5.3}$$

3. ω **(weight for private similarity)**: ω is the weight for how much the private similarity and the public similarity of OC' can be trusted if there are insufficient correspondent rating pairs between OC and OC'. Such a weight can be calculated based on the Chernoff Bound [114] as follows:

$$N_{min} = -\frac{1}{2\varepsilon^2} ln \frac{1-\gamma}{2}, \tag{5.4}$$

$$\omega = \begin{cases} \dfrac{N_{all}}{N_{min}}, & \text{if } N_{all} < N_{min}; \\ 1, & \text{otherwise}, \end{cases} \tag{5.5}$$

where ε is a small value (e.g., 0.1) representing a fixed maximal error bound which OC can accept, and $\gamma \in (0,1)$ is OC's confidence level about his/her own subjective assessments. In general, OC should trust OC''s private similarity more than OC''s public similarity since OC's own experience is more reliable. Thus, if the number of correspondent rating pairs between OC and OC' exceeds N_{min}, OC will only use $S_{pri}(OC \sim OC')$ in the calculation of OC''s relative trustworthiness.

4. $\overline{R_{TP}(OC')}$ **(average reputation of similar *TP*s with *OC'*)**: $\overline{R_{TP}(OC')}$ represents the weighted average of reputation values of TPs, the variation trends of whose objective assessments over time are similar to that of OC''s subjective assessments. Suppose there are m TPs, denoted as $\{TP_1, TP_2, \cdots, TP_j, \cdots, TP_m\}$, providing objective assessments for the l cloud services mentioned above. Following the time window partition method introduced above, we build *correspondent* assessment pairs between OC''s subjective assessments and TP_j's objective assessments for each cloud service, denoted as $(r_{OC',s_i}, oa_{TP_j,s_i})$, where oa denotes the value of objective assessments (e.g., $5ms$ for response time). All r_{OC',s_i} and oa_{TP_j,s_i} are then put together to build two assessment sequences ordered by the time of every time window, denoted as $\overrightarrow{r_{OC',s_i}}$ and $\overrightarrow{oa_{TP_j,s_i}}$ respectively. After that, each assessment sequence is converted into a ranking sequence according to the assessment values. If multiple assessments have the same value, their assessment ranking value is computed based on the average of their ranks. For example, if an assessment sequence expressed by fixed-number ratings is $\{4, 5, 2, 2, 3\}$ (a larger rating is the better), its ranking sequence is $\{2, 1, 4.5, 4.5, 3\}$. It should be noted that, if the nature of assessment values for a cloud performance aspect is "the smaller, the better" (e.g., response time), such values should be ranked in the ascending order, conversely in the descending order. Suppose the converted ranking sequences for $\overrightarrow{r_{OC',s_i}}$ and $\overrightarrow{oa_{TP_j,s_i}}$ are $\overrightarrow{x_{OC',s_i}}$ and $\overrightarrow{y_{TP_j,s_i}}$ respectively. Then, the similarity between these two ranking sequences are computed via Spearman's rank correlation coefficient [103], which is a common method to compute ranking similarity, as follows:

$$\rho(OC' \sim TP_j, s_i) = \frac{\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})}{\sqrt{\sum_{k=1}^{n}(x_k - \overline{x})^2 \sum_{k=1}^{n}(y_k - \overline{y})^2}}, \tag{5.6}$$

where n is the size of the ranking sequences, $1 \leqslant k \leqslant n$ indicates the corresponding positions in the sequences, and $\overline{x}$ and $\overline{y}$ are the average of the assessment ranking values in each ranking sequence. Hence, the average similarity of assessment variation trends between OC' and TP_j for all cloud services can be computed as follows:

$$\rho(OC' \sim TP_j) = \frac{1}{l}\sum_{i=1}^{l}\rho(OC' \sim TP_j, s_i). \tag{5.7}$$

All the TPs with $\rho(OC' \sim TP_j) > 0$ are then selected as the TPs whose objective assessments are similar to OC''s subjective assessments. Suppose there are p such TPs for OC', then the weighted average reputation of these TPs in Eq. (5.1) is computed as follows:

$$\overline{R_{TP}(OC')} = \frac{1}{p}(\sum_{q=1}^{p}\rho(OC' \sim TP_q) \times R_{TP_q}), \tag{5.8}$$

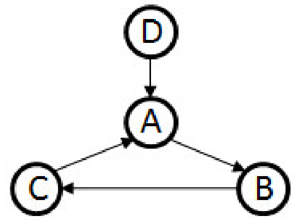


Figure 5.2: An Example of A Neighborhood Relationship

where R_{TP_q} represents TP_q's reputation on how truthfully its objective assessments are provided. The details of such reputations will be introduced in the next section.

5.1.2.2.2 Step 2. Computing Global Trustworthiness of *OC*s: Through Eq. (5.1), the authority center selects a fixed number of virtual neighbors for an OC according to the descending order of all other OCs' relative trustworthiness values, and maintains a virtual social network according to all these neighborhood relationships. If an OC is selected as a virtual neighbor of many other OCs, the OC should be considered trustworthy on providing truthful assessments.

In Zhang *et al.*'s work [189], the trustworthiness of a buyer is computed by counting the number of other buyers who select the buyer as their neighbor. However, such a method has a drawback in some cases. Consider a simple neighborhood relationship shown in Fig. 5.2. A's neighbor is B; B's is C; C's is A; D's is A. If we only count the number of their neighbors as their trustworthiness, A's trustworthiness is 2; B's, C's and D's trustworthiness are both 1. The maximum trustworthiness value in this example is 3. However, the trustworthiness of C and B should not be equal. That is because B is trusted by A, and C is trusted by B, but A's trustworthiness is higher than B's. Thus, B's trustworthiness should be slightly higher than C's. The similar opinion is supported by the literature [123, 172, 51], i.e., a party which is trusted by a more trustworthy party should be more trustworthy.

To this end, we apply the PageRank algorithm [123] in our model. Given a directed graph of neighborhood relationship G, and an OC is a vertex in G, then the *global* *trustworthiness* of the OC denoted as $Tr(OC)$ is computed as follows:

$$Tr(OC) = \frac{1-d}{N} + d \sum_{OC_i \in G(OC)}^{G(OC)} Tr(OC_i), \tag{5.9}$$

where $G(OC)$ is the set of all vertices who select the OC as their neighbor, N is the total number of vertexes in G and d is a damping factor which is commonly set to 0.85 [17] in the PageRank algorithm. In our model, $Tr(OC)$ is equivalent to the probability that a random OC' selects the OC as his/her neighbor, and d can be considered as a *trust transference degree* [171] between two OCs. In the example of Fig. 5.2, A's trustworthiness is 0.3326; B's is 0.3202; C's is 0.3097; D's is 0.0375. Finally, the *global trustworthiness* of every OC in the neighborhood relationship can be computed through the recursive Eq. (5.9).

5.1.3 The Sub Model for Computing Reputations of TPs

In the sub model for computing reputations of TPs, every TP offers objective assessments for the same cloud performance aspect assessed by OCs. The reputation of a TP depends on comparing its objective assessments to the majority of subjective assessments from OCs and the majority of objective assessments from other TPs. In the literature, many approaches [173, 118, 101] are proposed and validated to be effective for filtering out unfair ratings by comparing ratings with the majority opinion. Thus, our model follows this idea to determine whether a TP's behavior is credible.

We assume that there exists a conversion function[2], through which the values of objective assessments can be converted into normalized ratings introduced in Table 5.1. Suppose that, for a cloud service s_i, there is a sequence of normalized ratings, which is ordered by time and denoted as $\overrightarrow{r_{TP_j,s_i}}$, corresponding to the sequence of objective assessment values provided by a testing party TP_j. Then, $\overrightarrow{r_{TP_j,s_i}}$ is partitioned in the

[2]A simple way of defining such a function is to compare one objective assessment value of a cloud service for a performance aspect (e.g., $30ms$ for response time) with those of many other similar cloud services. After sufficient statistics, a reliable conversion function can be learned. Here, we apply the conversion function specified in our prior work [131].

Cases	Payoffs (TP_j)	$(r_{TP_j,s_i}, r_{\overline{TP},s_i})$	$(r_{TP_j,s_i}, r_{\overline{OC},s_i})$
1	ε_a	1	1
2	ε_b	1	0
3	ε_c	0	1
4	ε_d	0	0

Table 5.2: Reputation Payoff Matrix

same way of time window partition introduced above. In a time window, for s_i, there is at least one normalized objective rating r_{TP_j,s_i} from $\overrightarrow{r_{TP_j,s_i}}$, some subjective normalized ratings from OCs and some objective normalized ratings from other TPs. If there are multiple r_{TP_j,s_i} in a time window, then r_{TP_j,s_i} is the average of these ratings.

Let $r_{\overline{TP},s_i}$ denote the average of the objective ratings for s_i provided by all TPs except TP_j in a time window, and $r_{\overline{OC},s_i}$ denote the average of the subjective ratings provided by all OCs of s_i in a time window. In each time window, the authority center gives TP_j a reputation payoff to judge its behaviors in the time window. The reputation payoff matrix is illustrated in Table 5.2, where "1" means that the two corresponding ratings in a rating pair are in the same trust level, "0" means in different trust levels, and ε_a, ε_b, ε_c and ε_d are the reputation payoffs.

In a time window, the reputation payoff that TP_j can obtain depends on four cases as shown in Table 5.2.

Case 1: If r_{TP_j,s_i}, $r_{\overline{TP},s_i}$ and $r_{\overline{OC},s_i}$ are all in the same trust level, which means a high probability of TP_j providing truthful objective assessments of s_i.

Cases 2&3: If (r_{TP_j,s_i}, $r_{\overline{TP},s_i}$) or (r_{TP_j,s_i}, $r_{\overline{OC},s_i}$) are in the same trust level, but (r_{TP_j,s_i}, $r_{\overline{OC},s_i}$) or (r_{TP_j,s_i}, $r_{\overline{TP},s_i}$) are not, the probability of TP_j of providing truthful objective assessments should be lower than that in Case 1. Because objective assessments are usually considered more reliable than subjective assessments, the payoff in Case 2 should be higher than that in Case 3.

Case 4: If both $(r_{TP_j,s_i}, r_{\overline{TP},s_i})$ and $(r_{TP_j,s_i}, r_{\overline{OC},s_i})$ are all in the different trust levels, then TP_j is penalized by giving the least reputation payoff. The reputation payoffs can be defined in the inequality: $\varepsilon_a > \varepsilon_b > \varepsilon_c > \varepsilon_d > 0$.

Suppose that the total reputation payoffs that TP_j obtains by assessing s_i in the total t time windows are denoted as ξ_{TP_j,s_i}, then the reputation of TP_j based on s_i is computed as follows:

$$R_{TP_j,s_i} = \frac{\xi_{TP_j,s_i}}{t\varepsilon_a}, \tag{5.10}$$

and the reputation of TP_j for all cloud services is computed as follows:

$$R_{TP_j} = \frac{1}{l}\sum_{i=1}^{l} R_{TP_j,s_i}. \tag{5.11}$$

By introducing the majority of subjective assessments and the majority of objective assessments, our model can more accurately evaluate TPs' credibility of offering objective assessments.

5.2 Experiments

5.2.1 Experiment Setup

Because no suitable testing environment exists to evaluate our model. Like in the model evaluation in the related literature [100, 189], we simulate a cloud service environment based on our credibility evaluation framework. We collect the data of *response time* from CloudSleuth[3] for 59 real cloud services. The real data can describe the true variation trends of cloud service performance. To the best of our knowledge, there is no data set of subjective assessments published for the 59 cloud services. Hence, we select 8 cloud services having similar performance specifications from the 59 real cloud services. We then simulate subjective assessments from 300 OCs and objective assessments from 36 TPs for the *response time* of the 8 cloud services. Every OC consumes all the 8 cloud services and provides his/her subjective assessments,

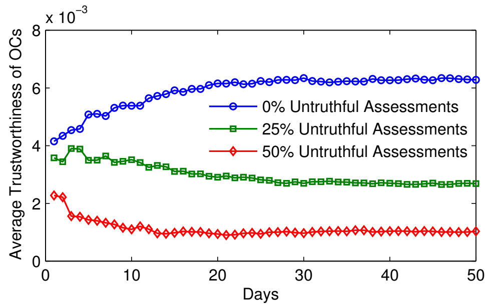
[3]www.cloudsleuth.net

and every TP provides objective assessments for every cloud service. We simulate the assessment behavior of all the participants in the cloud environment for a period of 50 simulated days. The trustworthiness of every OC and the reputation of every TP are computed and recorded at the end of each day. Each OC or TP has his/her/its own strategy on how truthful he/she/it provides assessments for the 8 cloud services or whether he/she/it is involved in a collusion attack.

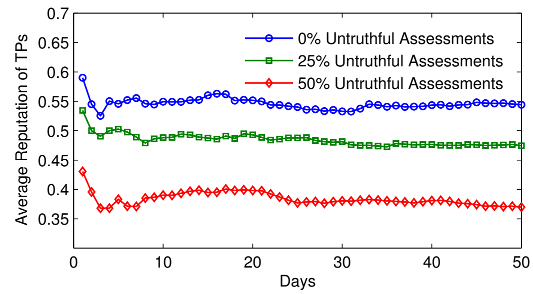
In our model, a collusion attack refers to that some users colluding to provide similar untruthful (too high or too low) assessments for a cloud service in order to manipulate the cloud service's reputation, and collusive assessments refer to such untruthful assessments. We require that each OC or TP has his/her/its own percentage of providing untruthful or collusive assessments. Here, untruthful assessments are randomly generated based on the real data of the 8 cloud services. An assessment is considered untruthful if it is in a different trust level with the corresponding real assessment data. In addition, considering subjective bias in subjective assessments, truthful subjective assessments in our experiments may have a small deviation compared to the corresponding real data of the truthful assessments.

5.2.2 Experimental Results without Collusion

In this set of experiments, some OCs or TPs randomly provide untruthful assessments without collusion. All the 300 OCs are equally divided into three groups. The OCs in each group provide $0\%, 25\%$ and 50% random untruthful subjective assessments respectively. Likewise, the 36 TPs are equally divided into three groups. The TPs in each group provide $0\%, 25\%$ and 50% random untruthful objective assessments respectively. For each OC, the authority center selects 10 other OCs as his/her virtual neighbors. The reputation payoffs $\{\varepsilon_a, \varepsilon_b, \varepsilon_c, \varepsilon_d\}$ is set to $\{1, \frac{2}{3}, \frac{1}{3}, 0\}$. The values of the reputation payoffs and the number of virtual neighbors can only affect the absolute values of TPs' reputations and OCs' trustworthiness, but cannot affect the tendency of our experimental results. Note that, the number of virtual neighbors should be far



(a) The Trustworthiness of OCs



(b) The Reputations of TPs

Figure 5.3: Experimental Results without Collusion

smaller than the total number of OCs. Because, for an OC, only a minority of other OCs who are considered the most trustworthy should be selected as the OC's neighbors.

Fig. 5.3 illustrates the average results of trustworthiness of OCs and reputations of TPs in every group over 50 days. The trustworthiness of an OC represents the probability that a random OC selects the OC as his/her neighbor. Hence, the value of the vertical axis in Fig. 5.3a is the average probability of an OC who is selected as a neighbor. The value of the vertical axis in Fig. 5.3b is the average reputation payoffs which a TP in each group can obtain. Fig. 5.3 shows that, for OCs, the more untruthful subjective assessments they provide, the lower the trustworthiness of the OCs, and for TPs likewise.

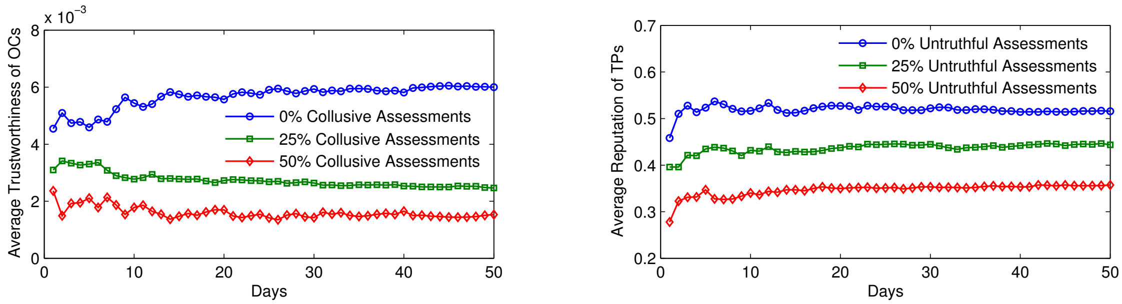
Next, we test the *tolerance* of untruthful assessments of our model, i.e, the maximum amount of untruthful assessments that our model can withstand to stay effective (i.e., the more untruthful assessments, the lower the trustworthiness of OCs or reputations of TPs). We fix the proportion of untruthful objective assessments from each group of TPs, and increase the proportion of untruthful subjective assessments from each group of OCs. Through 100 rounds of experiments, it is found that our model of evaluating cloud users' credibility can stay effective when the proportion of untruthful subjective assessments is smaller than approximately 55%. That's because even if half of the subjective assessments are untruthful, such untruthful assessments can still be detected by comparing the variation trends of the subjective assessments to those of objective assessments from TPs.

5.2.3 Experimental Results with Collusion

In this set of experiments, some OCs and/or TPs collude to provide similar untruthful assessments for the 8 cloud services. We follow the original setting of the experiments without collusion, i.e., each group of OCs or TPs provide $0\%, 25\%$ and 50% untruthful subjective or objective assessments. Three sub cases are considered in the experiments with collusion:

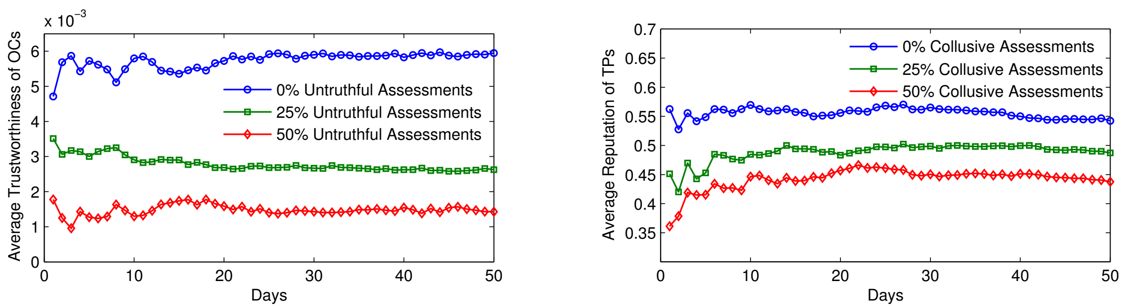
Sub Case 1: In this sub case, some OCs provide collusive assessments together, but some TPs still provide ordinary untruthful assessments according to their original strategy without collusion with the OCs. Fig. 5.4 illustrates the experimental results in this sub case, which demonstrates that the more the OCs provide collusive untruthful subjective assessments, the lower the trustworthiness of the OCs. Meanwhile, the tendency of the reputations of TPs is still preserved as that in the experiments without collusion. Like the tolerance testing in the experiments without collusion, we increase the proportion of the collusion subjective assessments in each group of OCs. The experimental results show that our model can stay effective when the proportion of collusive assessments is smaller than approximately 29%. This proportion is lower than that in the experiments without collusion, because a collusive OC can be easily selected as a neighbor of another collusive OC since they have similar assessments.

Sub Case 2: This sub case is the reverse of Sub Case 1: some TPs collude, but some OCs still provide ordinary untruthful assessments according to their original strategy without collusion with the TPs (i.e., 0%, 25% and 50% in each group respectively). The experimental results shown in Fig. 5.5 are similar to those in Sub Case 1, i.e., the more the TPs provide collusive untruthful objective assessments, the lower the reputations of the TPs. When the proportion of collusive objective assessments is smaller than approximately 40%, our model can stay effective. The reason



(a) The Trustworthiness of OCs (b) The Reputations of TPs

Figure 5.4: Experimental Results with Collusion (Sub Case 1)

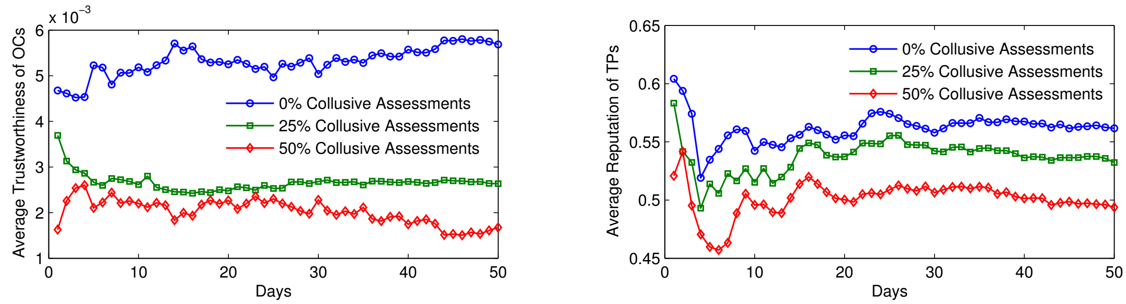


(a) The Trustworthiness of OCs (b) The Reputations of TPs

Figure 5.5: Experimental Results with Collusion (Sub Case 2)

for reaching such a high proportion to resist TPs' collusion is that, when a truthful objective assessment is quite different from the majority of objective assessments due to a collusion attack, such an assessment may be similar to the majority of subjective assessments, most of which are truthful. Thus, the reputation payoff is still paid in our model.

Sub Case 3: In this sub case, some OCs and TPs collude together. The proportions of collusive subjective assessments and objective assessments are set as the same as that in the original experimental setting (0%, 25% and 50%). The experimental results shown in Fig. 5.6 demonstrate the effectiveness of our model even if 25% of subjective assessments and objective assessments are collusive assessments. In the further experiments of tolerance of collusive assessments, it is found that our model can stay effective when the proportion of collusive assessments is smaller than



(a) The Trustworthiness of OCs (b) The Reputations of TPs

Figure 5.6: Experimental Results with Collusion (Sub Case 3)

approximately 26%.

In addition, we have conducted experiments in many different settings. As the experimental results are similar, we only introduce the main experimental results. Our experimental results demonstrate that our model can tolerate at least 25% collusive assessments in different situations. In the real world, reaching such a proportion for a collusion attack is highly costly. Hence, our model can effectively resist collusion in practice.

5.2.4 The Comparison of Untruthful/Collusive Assessment Tolerance

In this set of experiments, the capacities of untruthful or collusive assessment tolerance are compared among different models of evaluating users' credibility. Our proposed model is compared to Zhang *et al.*'s work [189] and the version of our proposed model without TPs, i.e., there are no TPs in this version, and only OCs' subjective assessments are used to compute their relative trustworthiness without objective assessments from TPs. The computation of global trustworthiness of OCs still follows the proposed approach. The initial proportions of untruthful or collusive assessments are set as the same as those in the experiments presented in the last section (i.e., 0%, 25% and 50%). Then, the proportions of untruthful or collusive subjective assessments are gradually increased to test the tolerance of the three compared models. Such tolerance

Subjective Assessments \ Models	Zhang *et al.*'s model [189]	Our model without TPs	Our model with TPs
Untruthful Assessments	30%	43%	55%
Collusive Assessments	21%	24%	29%

Table 5.3: Untruthful or Collusive Assessment Tolerance of Different Models

is assessed in two cases: the first case is that some OCs provide ordinary untruthful subjective assessments without collusion. The second is that some OCs provide collusive untruthful subjective assessments.

The experimental results are shown in Table 5.3, through which, we find that our model with/without TPs can achieve approximately 83%/43% improvement compared to Zhang *et al.*'s model in the case of providing untruthful assessments, and 38%/14% in the case of providing collusive assessments. Thus, evaluating cloud users' credibility based on both subjective assessments and objective assessments together is much more effective than that based on subjective assessments only in both the evaluation of users' credibility and the resistance of user collusion.

5.3 Conclusion

In this chapter, we have proposed a novel model for evaluating cloud users' credibility of providing subjective assessments or objective assessments for cloud services. Our model considers two different classes of cloud users (i.e., ordinary users and testing parties). The trustworthiness of *OC*s and the reputations of *TP*s are computed respectively to reflect how truthfully they provide subjective or objective assessments. Such trustworthiness and reputations can also influence each other, which means the model has the ability to resist user collusion to some extent. Hence, the cost of a successful collusion attack against our model would be high since a large number of cloud users need to participate. Sufficient experiments have been carried out under different circumstances according to cloud users' strategies. The experimental results have demonstrated that our proposed credibility evaluation model considering both subjective assessments and objective assessments significantly outperforms the existing work considering users' subjective assessment only. Thus, our model can be more effectively employed in cloud environments.

Chapter 6

CCCloud: Context-aware and Credible Cloud Service Selection

In Chapter 3, we have studied cloud service selection based on the comparison and aggregation of both subjective assessments and objective assessments. In Chapter 4, assessment contexts are taken into account in order to more accurately identify unreasonable assessments in terms of cloud consumers' personalized needs. In Chapter 5, we have studied how to evaluate cloud users' credibility of providing assessments, which is a crucial aspect impacting the effectiveness of cloud service selection.

In this chapter, we extend our proposed models and approaches in a more general circumstance, which considers cloud service selection in such a situation: cloud services are distributed worldwide. For each cloud service (*CS* for short), there are numerous ordinary cloud consumers (called Ordinary Consumers, *OC* for short) providing subjective assessments and a number of cloud monitoring and testing parties (called Testing Parties, *TP* for short) providing objective assessments for the cloud service. The assessments provided by these *CS*s, *OC*s and *TP*s have their own contexts. And the credibility of every *OC* or *TP* is evaluated to demonstrate how truthful his/her/its assessments are. The final result of cloud service selection is based on both subjective assessments and objective assessments according to cloud consumers' customized requirements and contexts.

To this end, this chapter proposes CCCloud: a credible and context-aware cloud service selection model based on subjective assessments from *OC*s and objective assessments from *TP*s. In this model, a novel credibility evaluation approach is proposed to detect biased or malicious cloud users who provide untruthful assessments, and can also resist user collusion. Moreover, our model considers subjective or objective assessments under different contexts, through which, the performances of cloud services can be more comprehensively and effectively evaluated from a potential consumer's perspective.

The features and contributions of CCCloud are summarized below:

- In contrast to most existing cloud service selection models, our model considers multiple performance attributes of cloud services. According to the characteristics of different attributes, those attributes can be assessed through subjective assessments, objective assessments, or a combination of both. Hence, our model based on both subjective assessments and objective assessments can comprehensively reflect the overall performance of cloud services.

- Our model takes objective assessments as benchmarks to filter out biased subjective assessments. Such a bias is usually caused by *OC*s' different preferences, contexts or malicious behavior in some cases. Thus, we assume that objective assessments without any subjective bias are considered more reliable than subjective assessments. The result of cloud service selection with less subjective bias more accurately reflects the real performance of cloud services.

- Our model takes assessment contexts into account based on two assessment features (i.e., *location* and *time*). When a potential cloud consumer requests cloud service selection, the context similarities between the consumer and different *TP*s are first computed to determine which *TP*(s) is/are more reliable. Then, all the *TP*s are grouped according to their contexts. Cloud service selection is carried out independently in every context group of *TP*s. In each context group, when the benchmark filtering is carried out, the context similarities between subjective assessments and objective assessments are computed to determine the benchmark levels, i.e., the more similar the contexts, the more reliable the subjective assessments, so that dynamic benchmark levels can be set to make such benchmark filtering more accurate.

- We propose an approach to evaluate the credibility of OCs and TPs providing subjective assessments and objective assessments in a context. An *OC* is considered more credible if his/her historical subjective assessments are more similar with the majority of subjective or objective assessments from *OC*s or *TP*s. In addition, the credibility of an *OC* can also be affected by the difference of variation trends between the *OC*'s subjective assessments and *TP*s' objective assessments over time. On the other hand, the credibility of a *TP* depends on the difference between its objective assessments and the majority of objective or subjective assessments from *TP*s or *OC*s. That implies that the credibility of *OC*s and *TP*s can be influenced by each other. That makes our model able to resist user collusion.

- We have conducted a series of experiments to evaluate the performance of our model under different circumstances, where some cloud users (OCs and/or TPs) behave maliciously according to different strategies. Compared to a well-known approach [101], our experimental results show that our model can more effectively evaluate user credibility, and enhance the resistance capability against user collusion. The final aggregated results of cloud service selection are computed based on the credibility and contexts of both OCs and TPs. Hence, the results delivered by our model can comprehensively and effectively reflect various potential consumers' preferences and customized requirements.

The rest of this chapter is organized as follows: the improved framework supporting CCCloud is introduced in Section 6.1. Section 6.2 presents the details of evaluating cloud users' credibility. Section 6.3 introduces the detailed processes of CCCloud. The experimental results are presented in Section 6.4

6.1 The CCCloud Framework

In this section, in order to support CCCloud, we present a new framework modified from our prior work introduced in Chapter 3. Figure 6.1 illustrates the framework, which is composed of four components, namely, (1) *cloud selection service*, (2) *benchmark testing management service*, (3) *user feedback management service*, and (4) *assessment aggregation service*.

- **Cloud Selection Service:** The *cloud selection service* is responsible for accepting and pre-processing the requests for cloud service selection from potential cloud consumers. In addition, it issues requests to the lower layer components. When a potential cloud consumer submits a request for selecting the most suitable cloud service, the *cloud selection service* firstly chooses those cloud services which can meet all the minimum quantitative functional or non-functional requirements (e.g., the type of services, targeted functions and costs) of the potential consumer from a candidate list of cloud services. Then, according to the consumer's further requirements, it sends requests to the *benchmark testing management service* and the *user feedback management service* for accessing the related records of alternative cloud services. These records are then sent to the *assessment aggregation service*, which returns the final aggregated score of each alternative cloud service to the *cloud selection service*.

- **Benchmark Testing Management Service:** The *benchmark testing management service* is responsible for collecting and managing objective assessments of cloud services from different *TP*s through benchmark monitoring and testing. In addition, it can request some *TP*s to carry out some specific cloud performance tests designed according to potential cloud consumers' requirements. Each monitored or tested performance aspect of a cloud service can be considered as an *objective attribute* of the cloud service. All these objective attributes are expressed in quantified forms (e.g., 90% for availability, $100ms$ for response time or 35.5 benchmark scores for CPU performance).

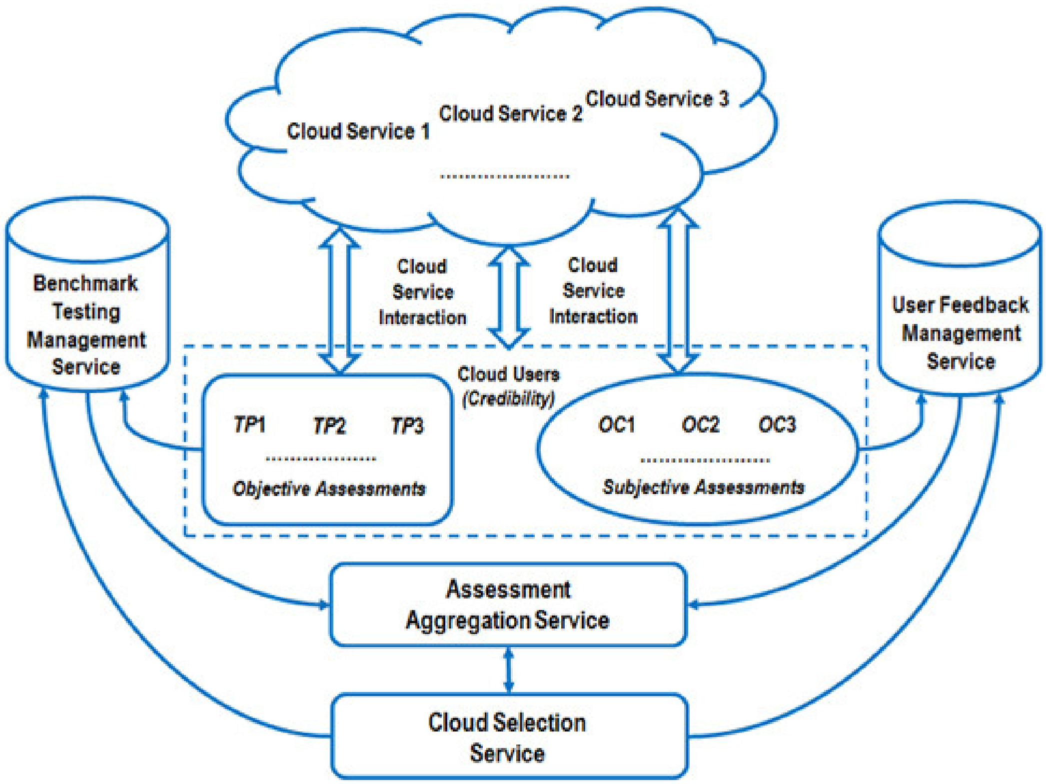


Figure 6.1: The Framework of CCCloud

- **User Feedback Management Service:** The *user feedback management service* is in charge of collecting and managing subjective assessments extracted from cloud consumer feedback. For each performance aspect of a cloud service, a consumer gives his/her subjective assessments in his/her context according to his/her experience and feelings. Each aspect that consumers assess can be considered as a *subjective attribute* of the cloud service. These subjective attributes are expressed by ratings (e.g., 1 - 5) or linguistic variables (e.g., “good”, “fair” and “poor”). Here, we follow the definition of *associated attributes* introduced in Chapter 3 and 4.

- **Assessment Aggregation Service:** The *assessment aggregation service* is responsible for processing assessments further and returning the final aggregated scores of every alternative cloud service to the *cloud selection service* according to potential cloud consumers' requirements. A potential consumer requesting for cloud service selection needs to specify under what context he/she will consume cloud services, and set the importance weights to every subjective or objective attribute. Through these weights, the potential consumer can also determine whether to put more trust on subjective assessments or objective assessments, so that the final scores of alternative cloud services can comprehensively reflect the various needs of different consumers.

6.2 The Credibility of Cloud Users

In this section, we present the details of our approach to the evaluation of the credibility of cloud users, *TP*s and *OC*s. The basic idea behind this approach is that the majority of assessments are taken as a "judge" to determine whether a cloud user's assessments are credible. In the literature, many approaches [118, 101] are proposed and validated to be effective for filtering out unfair ratings by comparing ratings with the majority opinion.

Specific to our approach, an *OC* having higher credibility means that his/her subjective assessments are more similar to the majority of subjective assessments from *OC*s and the majority of objective assessments from *TP*s; and the variation trend of his/her historical subjective assessments over time are more similar to that of the majority of objective assessments from *TP*s. Meanwhile, the majority of subjective or objective assessments should be computed through the subjective or objective assessments weighted by the corresponding *OC*s or *TP*s' credibility. Here, the reason of taking the similarity of assessment variation trends into account in our approach is that, though most of the cloud providers assure that the performances of their cloud services are consistent, a cloud service may not always perform consistently in practice. Thus, a cloud user's perception of a cloud service may vary with the real performance of the cloud service. In addition, objective assessments are considered more reliable than subjective assessments. If the variation trend of an *OC*'s subjective assessments is similar to that of the majority of objective assessments from *TP*s, such an *OC* should be considered honest, even though his/her subjective assessments are not similar to the majority of *TP*s' objective assessments. That is because, subjective assessments may contain personal subjective preferences. For example, suppose that an *OC* gives a subjective assessment sequence $\{1, 4, 2\}$ for a cloud service over time through a normalized rating system with the scale of $1 - 5$. Also suppose that, the sequence of the majority of *TP*s' objective assessments is $\{3, 5, 4\}$. Although the subjective assessments and the objective assessments are quite different, the variation trends of these assessments are very similar. It is possible that the *OC* may have subjective bias (e.g., giving lower or higher ratings according to their preference), but the similar variation trends indicate that such subjective assessments are very likely to be given honestly according to the *OC*'s own perception. Hence, only comparing the difference between an *OC*'s subjective assessments to the majority of assessments is not enough to reflect the *OC*'s honesty. In our model, we also use the similarity of the variation trends of assessments to adjust *OC*s' credibility.

On the other hand, a *TP* with higher credibility means that its objective assessments are more similar with the majority of all *TP*s' objective assessments and the majority of *OC*s' subjective assessments. Here, the variation trends of objective assessments are not taken to adjust *TP*s' credibility since objective assessments with scientific and statistical analysis should not contain subjective bias. For a cloud performance attribute, objective assessments should be only influenced by *TP*s' assessment contexts and honesty. Without any loss of generality, in our credibility evaluation approach, we require that objective assessments from different *TP*s are provided in similar contexts to avoid the situation where two assessments given by two credible *TP*s are quite different because their assessment contexts are quite different. In Section 6.3, we present how to aggregate objective assessments from *TP*s under different contexts for context-aware cloud service selection. In addition, we assume that subjective assessments and objective assessments can be normalized for a direct comparison in some way.

6.2.1 The Credibility of Ordinary Consumers

As introduced above, in CCCloud, the credibility of an *OC* can be affected by three factors. The first factor is the similarity between the variation trends of *OC*'s assessments for all subjective associated attributes and the majority of assessments for all corresponding objective attributes from *TP*s over time. The second factor is the difference between the *OC*'s assessments of all subjective attributes and the majority of those from all *OC*s. The last factor is the difference between the *OC*'s assessments for all subjective associated attributes and the majority of assessments for all corresponding objective attributes from *TP*s.

For a cloud service, let L_{OC} denote a non-empty set of *OC*s, and $|L_{OC}|$ denote the total number of *OC*s. These *OC*s provide subjective assessments to the cloud service over a period of time. All these assessments are ordered according to the time when they are submitted. Then, the assessment sequences are partitioned into mutually exclusive time windows. The length of each time window may be fixed or determined by the frequency of the submitted assessments. Moreover, it should be considerably small so that the performance of cloud services can hardly change in a time window. In a time window, there is/are one or more than one subjective assessments provided by each *OC* for the cloud service. The credibility of OC_i ($OC_i \in L_{OC}$) at the time window t is computed as follows:

$$\begin{aligned} Cr_t(OC_i) = Cr_{t-1}(OC_i) \times (1 + \rho_v \times F_v(A_i^a, M_o^a)) \times \\ [1 \pm F_s(A_i^s(t), M_s^s(t)) \pm F_o(A_i^a(t), M_o^a(t))], \end{aligned} \tag{6.1}$$

where $Cr_{t-1}(OC_i)$ denotes OC_i's credibility at the time window $t-1$. F_v, F_s and F_o represent the three factors to adjust OC_i's credibility: $F_v(A_i^a, M_o^a)$ denotes the factor on the similarity between the variation trends of OC_i's subjective assessments and the majority of TPs' objective assessments from the first time window to the time window t, where A_i^a denotes the OC_i's subjective assessment sequence over time and

M_o^a denotes the sequence of the majority of TPs' objective assessments over time. In addition, $\rho_v \in [0, 1]$ is an importance parameter to determine how fast $F_v(A_i^a, M_o^a)$ can influence an *OC*'s credibility. This parameter can be set by the potential consumer who requests for cloud service selection; $F_s(A_i^s(t), M_s^s(t))$ denotes the factor on the difference between OC_i's subjective assessments and the majority of subjective assessments from all OCs, where $A_i^s(t)$ denotes OC_i's subjective assessment for all subjective attributes at the time window t, and $M_s^s(t)$ denotes the majority of all OC's subjective assessments for all subjective attributes at t. Here, if OC_i submits more than one subjective assessment in a time window t, then $A_i^s(t)$ or $A_i^a(t)$ is the average of these subjective assessments. This setting can prevent a malicious OC_i from providing numerous untruthful assessments in a short period of time in order to manipulate cloud services' reputations. Similarly, $F_o(A_i^a(t), M_o^a(t))$ denotes the factor on the difference between OC_i's subjective assessment and the majority of TPs' objective assessments, where $A_i^a(t)$ denotes OC_i's subjective assessment for all subjective associated attributes at the time window t, and $M_o^a(t)$ denotes the majority of TPs' objective assessments for all corresponding objective associated attributes at t. The main notations in this section are summarized in Table 6.1.

Notations	Explanations
$Cr_t(OC_i)$	An ordinary consumer OC_i's credibility in the time window t
$Cr_t(TP_j)$	A testing party TP_j's credibility in the time window t
$F_v()$	The factor based on assessment **v**ariation trends (influencing OCs' credibility)
$F_s()$	The factor based on the majority of **s**ubjective assessments (influencing the credibility of OCs and TPs)
$F_o()$	The factor based on the majority of **o**bjective assessments (influencing the credibility of OCs and TPs)
$A_i^s(t)$	OC_i's subjective assessments for all **s**ubjective attributes
$A_i^a(t)$	OC_i's subjective assessments for all subjective **a**ssociated attributes
$M_s^s(t)$	The majority of all OCs' **s**ubjective assessments for all **s**ubjective attributes
$M_s^a(t)$	The majority of all OCs' **s**ubjective assessments for all subjective **a**ssociated attributes
$A_j^o(t)$	TP_j's **o**bjective assessments for all objective attributes
$A_j^a(t)$	TP_j's objective assessments for all objective **a**ssociated attributes
$M_o^o(t)$	The majority of all TPs' **o**bjective assessments for all **o**bjective attributes
$M_o^a(t)$	The majority of all TPs' **o**bjective assessments for all objective **a**ssociated attributes

Table 6.1: Notations in Section 6.2

The details of the three factors are introduced below:

1. The factor based on assessment variation trends: Suppose that there are n time windows in total, and the index k ($1 \leqslant k \leqslant n$) indicates the position of each time window. And there are s subjective attributes, o objective attributes and u associated attributes. In a time window t, the subjective assessment $A_i^s(t)$ is an s-element vector, in which each element corresponds to a subjective attribute. And the subjective assessment $A_i^a(t)$ is a u-element vector, in which each element corresponds to a subjective associated attribute. On the other hand, the majority of all OCs' subjective assessments $M_s^s(t)$ is an s-element vector. And the majority of TPs' objective assessments $M_o^a(t)$ is a u-element vector, in which each element corresponds to an objective associated attribute. Moreover, A_i^a or M_o^a is an assessment sequence composed of every $A_i^a(t)$ or $M_o^a(t)$ in every time window.

In a time window t, OC_i's subjective assessment $A_i^s(t)$ is the average of all his/her assessments in this time window. The majority of all OCs' subjective assessments $M_s^s(t)$ is the average of all OCs' subjective assessments weighted by their credibility in the time window $t-1$:

$$M_s^s(t) = \frac{\sum_i^{|L_{OC}|} A_i^s(t) \times Cr_{t-1}(OC_i)}{\sum_i^{|L_{OC}|} Cr_{t-1}(OC_i)}. \tag{6.2}$$

Likewise, the majority of TPs' objective assessments for all objective associated attributes $M_o^a(t)$ is computed as follows:

$$M_o^a(t) = \frac{\sum_j^{|L_{TP}|} A_j^a(t) \times Cr_{t-1}(TP_j)}{\sum_j^{|L_{TP}|} Cr_{t-1}(TP_j)}, \tag{6.3}$$

where L_{TP} is the set of TPs which test the same cloud service consumed by OCs, $Cr_{t-1}(TP_j)$ is the credibility of a TP_j in L_{TP} at t, and $A_j^a(t)$ is TP_j's normalized objective assessment, which is represented by a u-element vector, in which each element corresponds to an objective associated attribute.

In order to compare the variation trends between A_i^a and M_o^a, we first compare the variation trends between every pair of the corresponding elements (i.e., every pair of the corresponding associate attributes) in A_i^a and M_o^a. A_i^a and M_o^a are converted into ranking sequences according to their assessment values. If multiple assessments have the same value, their assessment ranking value is computed based on the average of their ranks (refer to Chapter 5).

Suppose that the converted ranking sequences for A_i^a and M_o^a are $\overrightarrow{x}$ and $\overrightarrow{y}$ respectively. Then, the similarity between these two ranking sequences are computed via Spearman's rank correlation coefficient [103], which is a common method to compute ranking similarity, as follows:

$$F_v(A_i^a, M_o^a) = \frac{\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})}{\sqrt{\sum_{k=1}^{n}(x_k - \overline{x})^2 \sum_{k=1}^{n}(y_k - \overline{y})^2}}, \tag{6.4}$$

where $\overline{x}$ and $\overline{y}$ are the average of the assessment ranking values in each ranking sequence respectively. Here, $F_v(A_i^a, M_o^a)$ is the similarity of variation trends for a single associate attribute pair. The overall $F_v(A_i^a, M_o^a)$ can be the average similarity for all the associated attribute pairs or computed based on the normalized importance weights set to every associate attribute pair by a potential cloud consumer who requests for cloud service selection. In the former condition, the credibility of OC_i can be influenced by the potential consumer's preference, i.e., which performance attributes can affect OCs' credibility more than others. This can promote the final service selection results generated from the potential consumer's perspective. The overall $F_v(A_i^a, M_o^a)$ varies between -1 and 1, which means that the two variation trends are the same and the exact opposite respectively.

2. The factor based on the majority of subjective assessments: As introduced above, $A_i^s(t)$ and $M_s^s(t)$ are both s-element vectors. $ED()$ denotes the Euclidean distance between two vectors. The factor $F_s(A_i^s(t), M_s^s(t))$ on the difference between the values of OC_i's subjective assessments and the values of the majority of OCs' subjective assessments for all subjective attributes is computed as follows:

$$F_s(A_i^s(t), M_s^s(t)) = (1 - \frac{ED(A_i^s(t), M_s^s(t))}{max(ED)}) \times \frac{f_s}{\rho_s}, \tag{6.5}$$

where $max(ED)$ represents the maximum Euclidean distance between two s-element vectors from a predefined assessment system (e.g., 1 - 5 rating system). ρ_s is an importance parameter (similar to the "pessimism factor" in [101]), which is used to determine how important the factor based on the majority of subjective assessments can influence OC's credibility, and how fast a dishonest OC's credibility can drop. ρ_s can be set according to the importance weights given by the potential consumer for all performance attributes. Its minimum value should be 2. And f_s represents the effect on adjusting an OC's credibility due to being similar or dissimilar to the majority of assessments. Here, we follow the idea in [101] to compute f_s as follows:

$$f_s = \begin{cases} 1 - \dfrac{ED(A_i^s(t), M_s^s(t))}{\sigma_s(t)}, & \text{if } ED(A_i^s(t), M_s^s(t)) < \sigma_s(t); \\ 1 - \dfrac{\sigma_s(t)}{ED(A_i^s(t), M_s^s(t))}, & \text{otherwise}, \end{cases} \tag{6.6}$$

where $\sigma_s(t)$ is the standard deviation of all *OC*'s subjective assessments in the time window t.

In Eq. (6.1), the sign before $F_s(A_i^s(t), M_s^s(t))$ depends on whether $A_i^s(t)$ and $M_s^s(t)$ are similar or not. If yes, OC_i is considered honest at t, thus his/her credibility on the factor based on the majority of subjective assessments will be increased and the sign should be set to $+$; otherwise, it is $-$, which means that OC_i should be punished since his/her subjective assessment is considered untruthful at t. If $A_i^s(t)$ and $M_s^s(t)$ are completely equal, $F_s(A_i^s(t), M_s^s(t))$ reaches the maximum.

3. The factor based on the majority of objective assessments: As we introduced above, subjective or objective assessments can be converted into normalized assessments (e.g., ratings), thus the factor based on the majority of objective assessments depends on the direct measurement of the difference between the values of OC_i's subjective assessments for all the subjective attributes and the values of the majority of TPs' objective assessments for all the corresponding objective attributes.

Therefore, similar to the computation of the second factor, $F_o(A_i^a(t), M_o^a(t))$ is computed as follows:

$$F_o(A_i^a(t), M_o^a(t)) = (1 - \frac{ED(A_i^a(t), M_o^a(t))}{max(ED)}) \times \frac{f_o}{\rho_o}, \tag{6.7}$$

and f_o is computed as follows:

$$f_o = \begin{cases} 1 - \dfrac{ED(A_i^a(t), M_o^a(t))}{\sigma_o(t)}, & \text{if } ED(A_i^a(t), M_o^a(t)) < \sigma_o(t); \\ 1 - \dfrac{\sigma_o(t)}{ED(A_i^a(t), M_o^a(t))}, & \text{otherwise}, \end{cases} \tag{6.8}$$

where $\sigma_o(t)$ is the standard deviation of all *TP*s' objective assessments in the time window t, and ρ_o is an importance parameter for the factor based on the majority of objective assessments. In general, ρ_o should not be larger than ρ_s since objective assessments are considered more reliable than subjective assessments. Likewise, the sign before $F_o(A_i^a(t), M_o^a(t))$ depends on whether $A_i^a(t)$ and $M_o^a(t)$ are similar.

6.2.2 The Credibility of Testing Parties

The computation of TPs' credibility is similar to that of OCs except that the factor based on assessment variation trends cannot influence TP's credibility because objective assessments do not generally contain subjective bias like subjective assessments. Thus, the computation of TPs' credibility can be computed as follows:

$$\begin{aligned} Cr_t(TP_j) = Cr_{t-1}(TP_j) \times \\ [1 \pm F_s(A_j^a(t), M_s^a(t)) \pm F_o(A_j^o(t), M_o^o(t))], \end{aligned} \tag{6.9}$$

where $A_j^a(t)$ is the normalized values of TP_j's objective assessment for all the objective associated attributes, and $M_s^a(t)$ is the normalized values of the majority of OCs' subjective assessments for all the corresponding subjective associated attributes at the time window t. $A_j^o(t)$ is the normalized values of TP_j's objective assessment for all the objective attributes, and $M_o^o(t)$ is the normalized values of the majority of TPs' objective assessments for all the objective attributes in the time window t. Both $A_j^a(t)$ and $M_s^a(t)$ are u-element vectors corresponding to all the associated attributes. And $A_j^o(t)$ and $M_o^o(t)$ are o-element vectors corresponding to all the objective attributes. Similarly, if TP_j gives more than one objective assessment at t, then $A_j^a(t)$ or $A_j^o(t)$ is the average of these objective assessments. Moreover, $M_o^a(t)$ and $M_o^o(t)$ in the time window t are computed through weighing every TP's credibility at $t-1$ (i.e., $Cr_{t-1}(TP)$) like Eqs. (6.2) and (6.3). At last, $F_s(A_j^a(t), M_s^a(t))$ and $F_o(A_j^o(t), M_o^o(t))$ are computed like Eqs. (6.5), (6.6) and (6.7), (6.8). If $A_j^a(t)$ and $M_s^a(t)$ are similar, the sign before $F_s(A_j^a(t), M_s^a(t))$ is $+$; otherwise $-$. Likewise, the sign before $F_o(A_j^o(t), M_o^o(t))$.

Note that, the initial credibility of an/a *OC* or *TP* can be any positive value. And if a new *OC* joins in at the time window t, his/her credibility is set to the average of all the current *OC*s' credibility, i.e., $\frac{\sum_i^{|L_{OC}|} Cr_t(OC_i)}{|L_{OC}|}$. Likewise, a new *TP*'s credibility is set to $\frac{\sum_j^{|L_{TP}|} Cr_t(TP_j)}{|L_{TP}|}$. In our approach, the credibility of OCs and TPs can be influenced by each other. OCs and TPs can be considered as "semi-trusted judges" for each other, which makes our approach more capable of resisting user collusion. Here, "semi-trusted" means that a party cannot be fully trusted, but can be trusted to some extent.

Distance Measurement between Multiple Ratings: in Eq. (6.1) and (6.9), we need an approach to determine whether two normalized assessments (i.e., ratings) are similar. Here, we adopt the approach proposed by Li and Wang [88] (refer to Section 5.1), which maps the *rating space* into a *trust space*, to measure the distance between two ratings.

6.3 Cloud Service Selection via CCCloud

In our prior models introduced in Chapters 3 and 4, subjective assessments and objective assessments are compared and aggregated through a modified fuzzy simple additive weighting system [26]. In this system, subjective assessments are represented using linguistic variables (e.g., “good”, “fair” and “bad”). That is because, in practice, people are used to expressing their opinions through human languages. Thus, it is customary for consumers to assess quality of services using linguistic variables. In the meantime, applying a fuzzy system can effectively deal with the inherent uncertainty of human languages. On the other hand, objective assessments in quantified forms can also be expressed through fuzzy numbers. Hence, in this chapter, we follow the fuzzy setting.

The detailed procedure of the cloud service selection via CCCloud consists of seven steps:

Step 1 (Set the importance weight for each attribute): According to a potential cloud consumer’s requirements, two importance weights in the form of linguistic variables are first set on how much to trust subjective assessments and objective assessments. Then, the potential consumer sets importance weights to all subjective attributes and objective attributes. All these weights are converted into trapezoidal fuzzy numbers (i.e., fuzzy weights) through a mapping (refer to Chapter 3).

Let $\widetilde{W_o}$ and $\widetilde{W_s}$ denote the fuzzy importance weights set for objective attributes and subjective attributes respectively, and $\widetilde{W_i}$ denote the fuzzy importance weight set for each subjective or objective attribute (there are s subjective attributes, o objective attributes and u associated attributes), where $i = 1, \cdots, s + o$. After that, W_i is the normalized weight of each attribute, which is computed as follows:

$$
\begin{aligned}
W_i &= \frac{d(\widetilde{W_s})}{d(\widetilde{W_s}) + d(\widetilde{W_o})} \times \frac{d(\widetilde{W_i})}{\sum_{i=1}^{s} d(\widetilde{W_i})}, \; i = 1, \cdots, s, \\
W_i &= \frac{d(\widetilde{W_o})}{d(\widetilde{W_s}) + d(\widetilde{W_o})} \times \frac{d(\widetilde{W_i})}{\sum_{i=s+1}^{s+o} d(\widetilde{W_i})}, \; i = s+1, \cdots, s+o,
\end{aligned}
\tag{6.10}
$$

where $\widetilde{W_s}$, $\widetilde{W_o}$ and $\widetilde{W_i}$ are expressed in trapezoidal fuzzy numbers. A trapezoidal fuzzy number denoted as $\widetilde{A} = (a, b, c, d)$, where $a < b < c < d$ are real numbers, belongs to a trapezoidal membership function. The most probable value of the evaluation data is represented in the interval $[b, c]$. The intervals $[a, b]$ and $[c, d]$ show the fuzziness of the evaluation data. Then, a defuzzification method is defined to convert fuzzy numbers into crisp numbers, i.e., the defuzzified value of $\widetilde{A}$ is its *signed distance*: $d(\widetilde{A}) = \frac{1}{4}(a + b + c + d)$ [26].

Step 2 (Group *TP*s according to their contexts): In our model, we assume that there are plenty of TPs distributed around the world. Here, we consider two assessment features (i.e., *location* and *time*) introduced in Chapter 4. In general, the *cloud selection service* should recommend the potential consumer which TPs should be selected to offer objective assessments for all alternative cloud services. The contexts of the selected TPs should not be quite different from the consumer's. In addition, the potential consumer can also select TPs according to his/her preference, and specify in what periods of time he/she hopes to consume cloud services.

Furthermore, all the selected TPs are divided into several groups according to their contexts. Suppose that there are l locations shown in all the selected TPs. As there are only two states for the assessment feature *time* in our model, i.e., *specified period of time* and *non-specified period of time*, all the TPs are divided into $2l$ groups. Note that, each TP can appear in two groups. For example, a TP in Sydney can appear in the context groups (*Sydney, specified time*) and (*Sydney, non-specified time*).

The procedure of cloud service selection for the consumer is first carried out independently in each context group. Thus, all objective assessments provided by a context group of TPs are given in the same context. However, in each group, the context of a subjective assessment can be different from that of objective assessments. For example, objective assessments from the TPs in the group of (*Sydney, specified time*) can be compared and aggregated with the subjective assessments with the context of (*Singapore, specified time*) in our model. Note that, the time specified in every assessment should be normalized into standard time due to time differences.

Suppose that the potential consumer wants to consume cloud service under the context C_p (e.g., (Sydney, 9am to 5pm)). Let C_q denote the context of a context group G_q, where $1 \leqslant q \leqslant 2l$. Through the proposed approach of context similarity computation introduced in Chapter 4, the similarity between C_p and each C_q is computed and denoted as $CSim(C_p, C_q)$.

Step 3 (Normalize subjective assessments into fuzzy ratings): Subjective assessments in our model are expressed by linguistic variables. Here, we apply a mapping from linguistic variables to trapezoidal fuzzy numbers (refer to Section 3.2). Every fuzzy number represents a fuzzy rating corresponding to a linguistic variable. In addition, a crisp rating corresponding to each linguistic variable is computed by the defuzzification method (i.e., *signed distance* in our model).

Step 4 (Normalize objective assessments into fuzzy ratings): In this step, we require a conversion function, through which the quantitative values of objective assessments can be normalized into the fuzzy ratings introduced in Step 2. An intuitive way of defining such a function is to compare one objective assessment value of a cloud service for a performance aspect (e.g., 30ms for response time) with those of many other similar cloud services. After sufficient statistics, a reliable conversion function can be learnt.

As trapezoidal fuzzy numbers can also express crisp values, e.g., 50 can be expressed by $(50, 50, 50, 50)$, objective assessments expressed by quantitative terms in our model are first expressed by fuzzy numbers. Then, these fuzzy numbers are converted into fuzzy ratings by comparing the values of the same objective attribute in all the alternative cloud services. For a TP in a context group, Let $X_i(j) = (a_i(j), b_i(j), c_i(j), d_i(j))$ denote the fuzzy value of the objective assessment for the ith objective attribute of the jth alternative cloud service. Then, the fuzzy rating denoted as $x_i(j)$ corresponding to $X_i(j)$ is computed as follows:

$$x_i(j) = \frac{X_i(j)}{\max_j(d_i(j))} \otimes 100, \text{or} \tag{6.11}$$

$$x_i(j) = \frac{\min_j(a_i(j))}{X_i(j)} \otimes 100, \tag{6.12}$$

where Eq. (6.11) is for the situation where the larger objective attribute value, the better, and Eq. (6.12) is for the situation where the smaller the objective attribute value, the better. $\max$ or $\min$ represents the maximum or minimum value in all the alternative cloud services. The details of the operations for fuzzy numbers can be found in Section 3.2.

Step 5 (Evaluate the credibility of *OC*s and *TP*s): In each context group, the credibility of *OC*s and *TP*s based on every alternative cloud service is computed through the approach introduced in Section 6.2. It should be noted that, the credibility evaluated in this step is not global, but customized based on the specific context of each context group. In different context groups, the same OC's credibility may be different. And according to the different potential consumers' requirements, the same TP's credibility may be different. In our model, the customized credibility instead of global credibility can promote a more effective result of cloud service selection for the potential consumer, since such credibility is specifically evaluated based on his/her perspective and customized requirements.

Step 6 (Filter biased subjective assessments): The process of filtering biased subjective assessments is carried out in every context group independently. Recall the definitions in Section 6.2, where in a time window for a context group, the values of the subjective associated attributes of every *OC*'s subjective assessment needs to be compared to the values of the corresponding objective associated attributes of the majority of the objective assessments. Following the notations in Section 6.2, $A_i^a(t)$ denotes OC_i's subjective assessment at time window t, and $M_o^a(t)$ denotes the majority of the objective assessments based on the specific context of their context group at t. Then, all $A_i^a(t)$ $(1 \leqslant i \leqslant |L_{OC}|)$ are grouped according to their contexts. Suppose there are l' locations shown in all the subjective assessments. Thus, there are at most $2l'$ groups of subjective assessments.

In each time window t of a context group, the Euclidean distance between $A_i^a(t)$

and $M_o^a(t)$ is computed. If $ED(A_i^a(t), M_o^a(t))$ exceeds a threshold $R_i(t)$, then $A_i^a(t)$ is considered quite biased and needs to be filtered out. Such a $R_i(t)$ is computed as follows:

$$\begin{aligned} R_i(t) = (1 - &\frac{d(\widetilde{W}_o)}{d(\widetilde{W}_s) + d(\widetilde{W}_o)}) \times \\ &\frac{CSim[cx(A_i^a(t)), cx(M_o^a(t))]}{CSim[cx(M_o^a(t)), cx(M_o^a(t))]} \times max(ED), \end{aligned} \tag{6.13}$$

where $cx(A_i^a(t))$ and $cx(M_o^a(t))$ denote the contexts of $A_i^a(t)$ and $M_o^a(t)$ respectively. Here, in order to offset the effect caused by the constant C in the modified bipartite SimRank algorithm introduced in Section 4.1, $CSim[cx(M_o^a(t)), cx(M_o^a(t))]$ is applied to represent the similarity between the contexts of $M_o^a(t)$ and itself. From Eq. (6.13), when the potential consumer trusts objective assessments more, $R_i(t)$ becomes smaller, so that more subjective assessments are considered biased and will be filtered out. In addition, when the context similarity $CSim[cx(A_i^a(t)), cx(M_o^a(t))]$ becomes lower, $R_i(t)$ becomes smaller. That means the subjective assessments are given in a more different situation compared to the objective assessments, thus such subjective assessments are considered less reliable and need to be filtered out more rigorously.

Step 7 (Aggregate assessments): In each context group, after the filtering process in Step 6, subjective assessments and objective assessments are aggregated according to OCs and TPs' credibility evaluated in Step 5. In the time window t, suppose that there are $|L'_{OC}|$ OCs left after filtering, and $OC_{i'}$ denotes one of these OCs. Let L_{CS} denote the set of all the alternative cloud services. The overall assessment $M_k(t)$ for an alternative cloud service $CS_k \in L_{CS}$ in the time window t is computed as follows:

$$\begin{aligned} M_s^s(t) &= \frac{\sum_{i'}^{|L'_{OC}|} A_{i'}^s(t) \times Cr_t(OC_{i'})}{\sum_{i}^{|L'_{OC}|} Cr_t(OC_{i'})}, \\ M_o^o(t) &= \frac{\sum_{j}^{|L_{TP}|} A_j^o(t) \times Cr_t(TP_j)}{\sum_{j}^{|L_{TP}|} Cr_t(TP_j)}, \\ M_k(t) &= M_s^s(t) \oplus M_o^o(t), \end{aligned} \tag{6.14}$$

where $M_s^s(t)$ is an s-element vector corresponding to subjective attributes, and $M_o^o(t)$ is an o-element vector corresponding to objective attributes. The operator $\oplus$ represents *concatenation*, thus $M_k(t)$ is a $(s+o)$-element vector corresponding to all attributes.

After that, $M_k(t)$ is weighted by the normalized importance weights computed in Eq. (6.10). Let $\overline{M_k^q(t)}$ denote the weighted score for the alternative cloud service CS_k in the time window t in the context group G_q. Then, the final score of CS_k denoted as $F_k(t)$ is weighted by the context similarity between the potential consumer and every context group:

$$F_k(t) = \frac{\sum_q^{2l} \overline{M_k^q(t)} \times CSim(C_p, C_q)}{\sum_q^{2l} CSim(C_p, C_q)}. \tag{6.15}$$

Finally, according to the final scores of all the alternative cloud services, the cloud services are ranked for selection. Through our model, all the alternative cloud services can be effectively and comprehensively evaluated based on the potential consumer's perspective and customized requirements. In addition, besides the fuzzy rating system applied above, our model can also be suitable for any other multiple rating systems in practice.

6.4 Experimental Evaluation

6.4.1 Experiment Setup

Since there is no suitable real data supporting our cloud service selection model, we simulate a cloud service environment based on the proposed framework introduced in Section 6.1. The data used in this environment is partially collected from real cloud services, and partially generated synthetically based on real cloud services.

In our experiments, there are three subjective attributes: *privacy* (a_1), *after-sales services* (a_2), *response time* (a_3); and two objective attributes: *response time* (a_4) and *CPU performance* (a_5), where a_3 and a_4 are the associated attribute pair. We collect the data of *response time* (a_4) from CloudSleuth and the data of benchmark scores of *CPU performance* (a_5) from CloudHarmony for 59 real cloud services. To the best of our knowledge, there is no published dataset of subjective assessments from ordinary cloud consumers for the 59 real cloud services, thus, we synthetically generate subjective assessments from 300 OCs. We select 10 cloud services having similar performance specifications from the 59 real cloud services, and then synthetically generate objective assessments from 48 TPs, which are equally divided into two groups to simulate different context groups of TPs. The TPs in each group have similar contexts. As the collected real data can describe the real variation trends of cloud service performance, we generate truthful subjective assessments for a_3 according to the ranking of the real data of *response time* in a_4, and subjective assessments of a_1 and a_2 are randomly generated. In addition, through the real data, objective assessments are normalized for a direct comparison with subjective assessments.

In our experiments, we require that every OC consumes all the 10 alternative cloud services and provides his/her subjective assessments, and every TP also provides objective assessments for every cloud service. We simulate the assessment behavior of all the cloud users for a period of 30 simulated days. The credibility of the OCs and TPs are computed and recorded at the end of each day. And each OC or TP has his/her/its own strategy on how truthful he/she/it provides assessments or whether he/she/it is involved in a collusion attack. Here, a collusion attack refers to the case that some malicious users intentionally provide similar untruthful (too high or too low) assessments for a cloud service, and collusive assessments refer to such untruthful assessments. We require that each OC or TP has his/her/its own percentage of providing untruthful or collusive assessments. Untruthful assessments are randomly generated based on the real data of the 10 cloud services, i.e., an assessment is considered untruthful if it is in a different trust level with the corresponding real assessment data. In addition, considering subjective bias in subjective assessments, truthful subjective assessments in our experiments may have a small deviation compared to the corresponding real data of the truthful assessments.

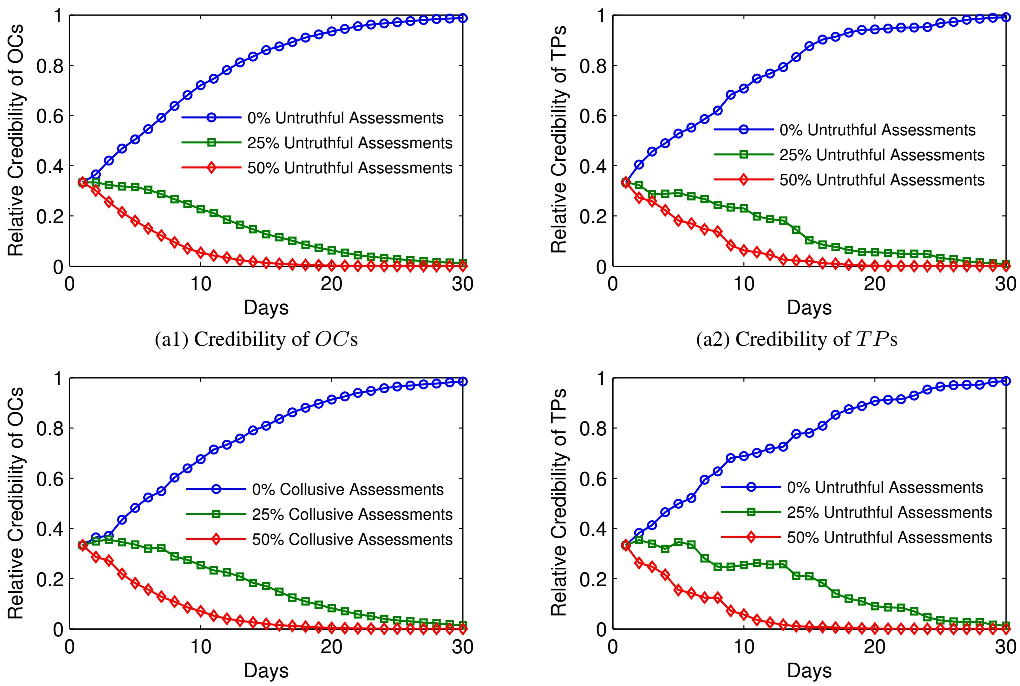
6.4.2 Experiment 1

In this experiment, we validate our proposed credibility evaluation approach introduced in Section 6.2. In each context group, all OCs or TPs are divided into three groups. The OCs or TPs in each group provide different percentages of random untruthful or collusive assessments. The default percentages for each group are 0%, 25% and 50% respectively. We have conducted experiments in two cases:

Case 1: some OCs and TPs provide random untruthful assessments independently;

Case 2: some OCs provide collusive assessments, but some TPs still provide random untruthful assessments without collusion

Here, we assume that TPs should be hardly involved in collusion with ***OC***s, but can provide random untruthful assessments, since TPs are considered semi-trusted in our approach. Fig. 6.2 illustrates the experimental results of the two cases. The horizontal axes are the 30 simulated days. And vertical axes are the relative credibility of ***OC***s or ***TP***s, i.e., the values of the relative credibility of OC_i or TP_j at the vertical axes are $\frac{Cr_t(OC_i)}{\sum_i^{|L_{OC}|} Cr_t(OC_i)}$ or $\frac{Cr_t(TP_j)}{\sum_j^{|L_{TP}|} Cr_t(TP_j)}$. The relative credibility of an/a ***OC***/***TP*** is his/her/its credibility over the sum of the credibility of all ***OC***s/***TP***s. The trend of the experimental results in the two cases is similar. The credibility of OCs/TPs exhibits some random fluctuations in the first few days, but their credibility returns to and maintain the normal trend in the following days. That is because, our credibility evaluation approach needs some historical assessment records to adjust OCs or TPs' credibility. Fig. 6.2 demonstrates that the more collusive/random untruthful assessments the OCs/TPs give, the lower the relative credibility of the OCs/TPs, thus our approach can effectively detect the OCs or TPs who/which provide untruthful or collusive assessments.



(a1) Credibility of OCs (a2) Credibility of TPs

(b1) Credibility of OCs (b2) Credibility of TPs

Figure 6.2: Experimental Results for Credibility Evaluation

6.4.3 Experiment 2

In this experiment, we evaluate our credibility evaluation approach on estimating the real performance of a cloud service. We follow the setting in Experiment 1, i.e., the default percentages of untruthful/collusive assessments from OCs are set to 0%, 25% and 50%. Furthermore, we compare our approach to a well-known approach - RATEWeb [101], through which, a rater's credibility can be evaluated in service-oriented environments. In [101], the reputation score of a service is computed via all raters' ratings weighted by every rater's credibility. A rater is considered credible if its ratings are similar with the majority of ratings and the current reputation score of a service it consumes.

Fig. 6.3 illustrates the experimental results in four cases, i.e., cloud services perform consistently or inconsistently, and some OCs provide random untruthful assessments or collusive assessments. In Fig. 6.3, the left subfigures show the estimated scores of a cloud service computed via our approach and RATEWeb, and the real performance scores of the same cloud service over 30 days. And the right subfigures show the differences between the estimated scores via our approach and RATEWeb and the real scores. In Fig. 6.3 (a), the cloud service performs consistently, and some OCs just provide untruthful assessments. In this case, the performance of our approach and that of RATEWeb are quite similar. However, in Fig. 6.3 (b), our approach performs better than RATEWeb, i.e., in the most of days, the score differences based on our approach are smaller than those based on RATEWeb. That is because, RATEWeb assumes that a service performs consistently, thus, in RATEWeb, a rater's credibility can be affected by the difference between its current ratings and the current reputation of the service. However, our approach does not have this assumption. Furthermore, Fig 6.3 (c)$\&$(d) shows that our approach performs better than RATEWeb whether or not a cloud service performs consistently. That is because, in our approach, TPs are taken as "semi-trusted judges" to adjust OC's credibility, but there is no such a mechanism in RATEWeb. Thus, our model performs better in the collusion situation. In this case, our approach achieves approximately 34.5% improvement compared to RATEWeb on reducing the difference between estimated scores and real performance.

Moveover, we compare our approach to RATEWeb and our prior work [134] introduced in Chapter 4, in which OCs and TPs' credibility is not taken into account, in the situations of different percentages of untruthful/collusive assessments. Here, we take the *success rate* as a metric. If an estimated score and the real performance score of a cloud service are *similar* (i.e., in the same trust level), then such an estimate is considered successful. The experimental results in Fig. 6.4 illustrate that, with the increasing percentages of untruthful/collusive assessments, our approach always performs better than RATEWeb and our prior work. Note that, in the collusion situation, when the percentage of collusive assessments is small (about 20%), the success rates based on our approach and RATEWeb are very close (about 85%). With the increasing percentages of collusive assessments, the success rates based on our approach and RATEWeb all drop, but the success rate of RATEWeb drops more dramatically. The minimum success rate based on our model stays at about 55%, and that based on RATEWeb drops to under 10%. That is because, if most of OCs are involved in collusion, such collusive OCs would be dominant to manipulate the credibility evaluation via RATEWeb. Then, the credibility evaluation error will be further enlarged by the credibility system in RATEWeb. In that case, our prior work [134] without consideration of credibility outperforms RATEWeb due to the lack of such credibility errors.

6.4.4 Experiment 3

In this experiment, we evaluate the overall performance of CCCloud. We equally divide all the 300 OCs and 48 TPs in two groups. Then, we set that one group of 150 OCs and 24 TPs are in *Sydney*, and the other group of the rest are in *Hong Kong*. We assume that a potential consumer wants to select a cloud service from the 10 alternative cloud services, and consume it under the context (*Sydney, Morning*).

We first generate truthful objective assessments in the Sydney group. Compared to the objective assessments in the Sydney group, the objective assessments in the Hong Kong group for the same cloud services may contain some bias to simulate the truthful objective assessments under quite a different context. Here, a bias level denoted as BL is set to represent how much the biased ratings deviate from the normal synthetic ratings, where $BL = 4, \cdots, 8$ since a fuzzy rating scale of 1-9 is employed in our model. The conditions that $BL = 1, \cdots, 3$ are not considered here since such biases are too small to reflect the effect of assessment differences caused by different contexts in practice. Then, we generate untruthful or collusive subjective assessments in the two groups. After that, we carry out cloud service selection via our model and RATEWeb in the two groups of cloud users, where our model takes contextual subjective assessments and objective assessments into account, but RATEWeb only considers subjective assessments without contexts. Here, the setting of the percentages of untruthful or collusive assessments follows Experiment 1.

In the results, two ranks of the alternative cloud services, denoted as R_c and R_r,

With/Without Collusion	Models \ BL	4	5	6	7	8
Without Collusion	RATEWeb	0.9287	0.9195	0.9130	0.8892	0.8472
	CCCloud	**0.9294**	**0.9336**	**0.9315**	**0.9483**	**0.9407**
With Collusion	RATEWeb	0.8996	0.9024	0.8823	0.8553	0.8212
	CCCloud	**0.9042**	**0.9182**	**0.9119**	**0.9278**	**0.9260**

Table 6.2: Accuracy Comparison based on Ranking Similarity (Experiment 3)

are obtained through our model and RATEWeb respectively. In addition, a rank of the alternative cloud services, denoted as R_o, is computed based on the original assessments without untruthful or collusive assessments. As R_o is computed without any noise, R_o is taken as the metric rank. Let $R_{sim}()$ denote the ranking similarity calculation through the *Kendall tau rank distance* [36] which is a common metric to measure the distance between two ranks through counting the number of pairwise disagreements between the two rankings. If $R_{sim}(R_o, R_c) > R_{sim}(R_o, R_r)$, that means our model is more effective than RATEWeb. The ranking similarity results in Table 6.2 show that our model can more accurately select the suitable cloud service than RATEWeb from the potential consumer's perspective in both the collusion and no collusion situations with different BLs. Every value in Table 6.2 is the average ranking similarity computed based on 50 rounds of experiments, so that the generality of experimental data can be kept in our experiments. In general, compared to RATEWeb, our model can achieve approximately 10% improvement at most. Note that, though our credibility evaluation approach significantly outperforms RATEWeb (in Experiment 2) in estimating the performances of cloud services, the results in Experiment 3 illustrates that RATEWeb is an effective model for ranking alternative cloud services, but CCCloud performs better than RATEWeb in all the situations. Then, we carry on many more rounds of experiments. In 98% of these experiments, our model outperforms RATEWeb. The few exceptional cases occur due to the noise from the random generation of the synthetic experimental data in these cases.

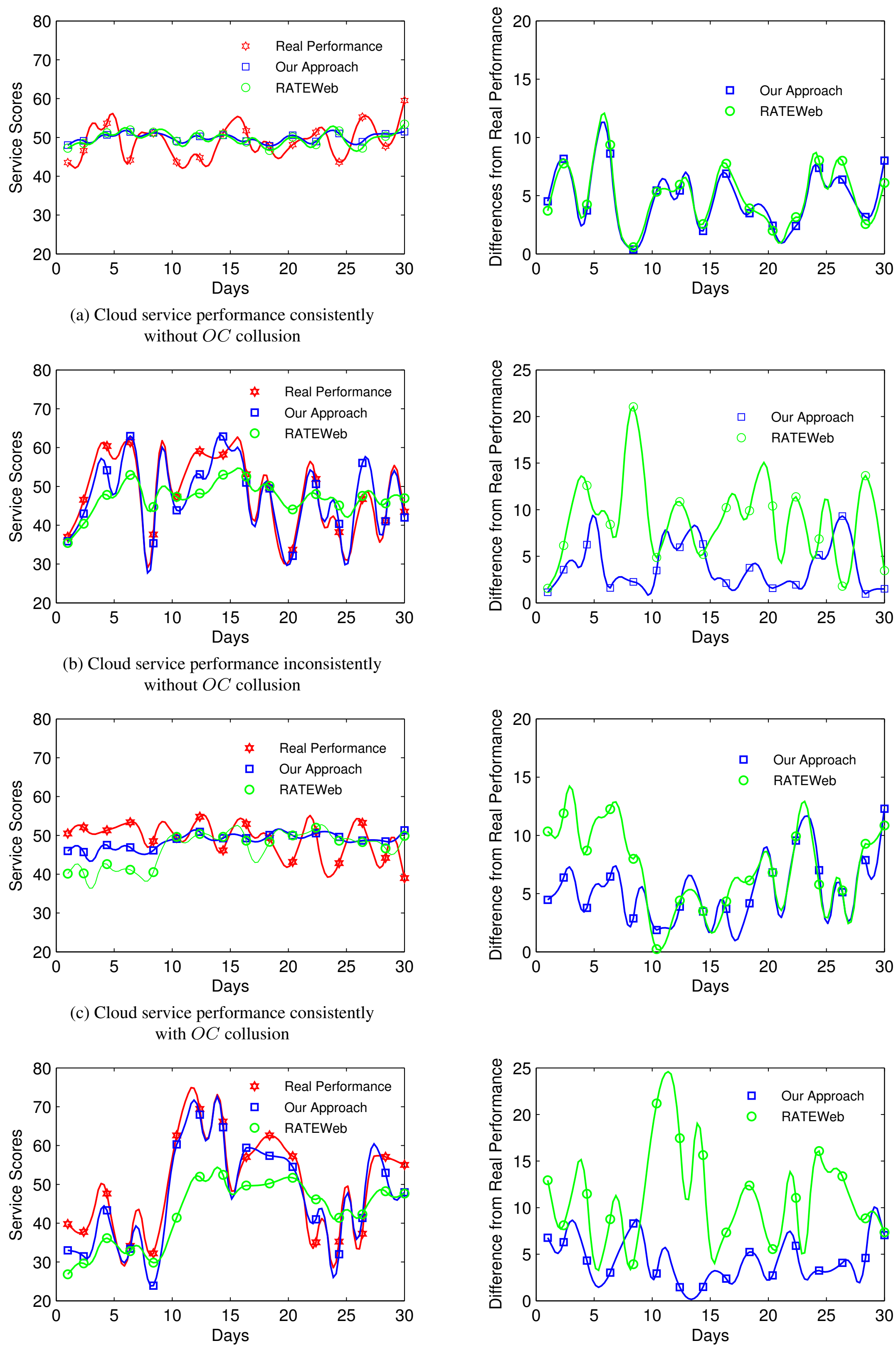
6.5 Conclusion

In this chapter, we have proposed CCCloud: a credible and context-aware cloud service selection model based on the comparison and aggregation of subjective assessments extracted from ordinary cloud consumers and objective assessments from professional performance monitoring and testing parties.

We have first proposed a novel approach to evaluate the credibility of both ordinary consumers (*OC*s) and testing parties (*TP*s), where the majority of assessments from both *OC*s and *TP*s are used to adjust their credibility. In addition, the variation trends of assessments are also used to adjust *OC*s' credibility, so that *OC*s' subjective preferences are considered in the credibility evaluation. In this approach, the credibility of *OC*s and *TP*s can be influenced by each other, which makes that our approach not only accurately evaluates how truthfully they assess cloud services, but also resists user collusion.

Then, we have considered assessment contexts in cloud service selection. The context similarity between subjective assessments and objective assessments is used to dynamically adjust benchmark levels which are used to filter out biased subjective assessments. In addition, the context similarity between a potential consumer and different *TP*s is computed to weight the results of cloud service selection generated from *TP*s in different contexts. The final results of cloud service selection based on assessment credibility are comprehensive and customized according to the potential consumer's requirements and context.

We have conducted experiments in many different settings. The experimental results demonstrate that CCCloud outperforms the existing models under different situations, especially in the situations where the performance of a service is inconsistent, and malicious cloud users give collusive assessments.



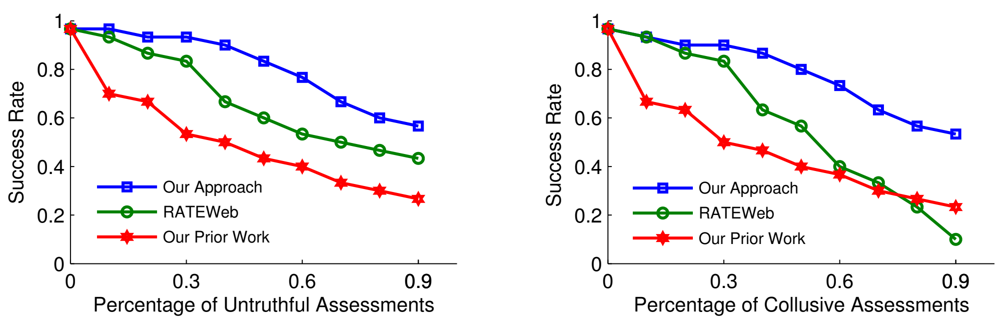
(a) Cloud service performance consistently without OC collusion

(b) Cloud service performance inconsistently without OC collusion

(c) Cloud service performance consistently with OC collusion

(d) Cloud service performance inconsistently with OC collusion

Figure 6.3: Estimated Performances Compared to Real Performances



(a) Without Collusion (b) With Collusion

Figure 6.4: Success Rates in Different Situations

Chapter 7

An Incentive Mechanism for Eliciting Continual and Truthful Assessments

As introduced in Chapter 3, cloud services are typically evaluated based on subjective assessments from ordinary consumers and objective assessments through quantitative performance monitoring and testing carried out by professional testing parties[1]. No matter what types of assessments are applied, the credibility of cloud users' assessments has a great impact on the accuracy of cloud service evaluation and selection. The traditional way of improving selection accuracy is to evaluate users' credibility in order to find out more effective assessments. In this chapter, we consider this problem from a novel perspective, i.e., motivating cloud users to actively provide truthful assessments, and thus improving selection accuracy.

In cloud environments, service performance may vary substantially and frequently due to the dynamic nature of cloud services. Thus, continual assessments over time are needed to effectively reflect the dynamic performance of services. However, eliciting continual and truthful assessments in cloud environments is still a challenging problem since it is usually hard to make self-interested users behave cooperatively in an online community [12]. A cloud user usually does not have sufficient incentives to regularly provide assessments of cloud services on time. To motivate such users, an effective incentive mechanism should be designed. A common solution is that a cloud user can

[1] For avoiding ambiguity, ordinary consumers and testing parties are called *cloud users* in this chapter.

be paid if it provides assessments on schedule. The monetary rewards[2] may be provided by some professional cloud evaluation organizations, such as CloudReviews[3], the aim of which is to provide cloud selection services to potential cloud consumers based on cloud users' assessments and therefore earn profits from the potential consumers. However, such a simple mechanism cannot prevent a user from "free-riding" (i.e., providing arbitrary assessments) [97, 193]. Moreover, sometimes an honest user could also provide arbitrary assessments in order to obtain monetary rewards when it does not really know the real performance of cloud services (e.g., a user does not consume services on the scheduled time while a user is required to provide an assessment). Such arbitrary assessments may be erroneous and misleading, and therefore greatly affect the effectiveness of service evaluations. To avoid the submission of such arbitrary assessments, an effective incentive mechanism should motivate users to always tell the truth, i.e., allowing honest users to provide uncertain assessments to express their uncertainty about service performance when necessary.

In order to design an uncertain assessment compatible incentive mechanism, the process of cloud users providing assessments should be theoretically modeled first. Then the possible strategies of users need to be analyzed. In order to be compatible with uncertain assessments, a novel and effective incentive mechanism should be designed based on users' strategies. Moreover, an optimal incentive mechanism needs to be further designed in order to achieve some specific goals according to different situations, e.g., maximize users' rewards or the total benefits. However, in the literature, the study of designing such an incentive mechanism is still missing.

Different from all the prior incentive mechanisms, which do not consider uncertain assessments, in this chapter, we propose an uncertain assessment compatible incentive mechanism for eliciting continual and truthful assessments of cloud services. The features and contributions of our work are summarized as follows:

(1) Under our proposed mechanism, a user is considered "honest" if it gives truthful

[2] The rewards can be paid in any form, e.g., points, discount and privileges, which can be taken as monetary rewards.

[3] www.cloudreviews.com

assessments most of the time, but may give a small number of uncertain assessments once it is not sure about the real performance of a service. The word "honest" indicates such a user always tells the truth. Thus, an UAC (*uncertain-assessment-compatible*) assessment scheme is first proposed, which can be extended from any type of ordinary (subjective or objective) assessment system, but includes an extra uncertain state (see Section 7.1.1). Then the behaviors of users providing assessments are modeled using a repeated game framework (see Section 7.1.2).

(2) A user can receive monetary rewards from a professional organization (called a *broker*) mentioned above for regularly providing assessments on schedule via a user agent for the cloud services it consumes. We propose an assessment scoring scheme for controlling the monetary rewards (see Section 7.1.3). In a nutshell, truthful assessments would bring the most rewards; uncertain assessments would bring less rewards; untruthful or arbitrary assessments would bring the very least rewards. Through our proposed mechanism, a rational user would choose his/her best option, i.e., providing truthful assessments. Once it is not sure about service performance, there still exists a second-best option, i.e., providing uncertain assessments.

(3) In order to build an effective incentive mechanism, we present the theoretical analysis (see Sections 7.1.4) of the scoring scheme according to the different strategies of users (i.e., providing truthful/uncertain/untruthful/arbitrary assessments). Moreover, we discuss how to build an optimal incentive mechanism in our scenario (see Section 7.1.5) and the feasibility of solving the *whitewashing* problem [41] based on our proposed mechanism (see Section 7.1.6).

(4) The results from the theoretical analysis show that our approach is effective in most circumstances (see Section 7.2.1). Furthermore, in order to evaluate the practical feasibility of our approach, we carry out simulation experiments under different situations. The results from the simulation strongly support the theoretic analysis (see Section 7.2.2).

7.1 Incentive Mechanism Design

The basic idea behind our approach is as follows: cloud users can get paid by selling their assessments for cloud services to a *broker* via a user agent system. Cloud users are allowed to provide uncertain assessments for the services when they are not sure about the real performance of the services. The cloud performance evaluation is carried out by the broker based on cloud users' assessments, and the broker pays monetary rewards to the current cloud users for their assessments and obtains profits from potential cloud consumers by offering cloud selection services.

A user's incentive is represented through its expected long-term payment. The long-term payment is composed of the payments obtained in the continual time windows, e.g., 9am - 10am every day. Through an assessment scoring scheme, users' participation of selling their assessments are controlled. In brief, if a user is considered to submit a truthful assessment in a time window, it can keep on selling assessments until it is considered to have submitted an uncertain or untruthful assessment in a subsequent time window. Due to the submitted uncertain or untruthful assessment, the user would be isolated from selling assessments for a period of time, so that its long-term payment would suffer a loss because of such isolation. This is like fixed-term imprisonment. After the "imprisonment", the user can still be involved in the subsequent assessment transactions. Hence, in a time window, the user would believe that truthful reporting can maximize its long-term payoff and an uncertain assessment would bring a larger payoff than an untruthful or arbitrary one, if the broker can correctly judge the truthfulness of an assessment with an overwhelming probability.

7.1.1 The UAC Assessment Schemes

A cloud user can give its own assessments for different performance aspects of cloud services it consumes. For each aspect, such assessments can be expressed in any reasonable form including subjective or objective assessments. Taking service response time as an example, a cloud user can give its numerical ratings (e.g., "1", "2" or "3") or linguistic ratings (e.g., "poor", "fair" or "good") to express its subjective assessments. On the other hand, a user can also provide objective assessments according to QoS testing (e.g., 200ms for response time). For any type of an assessment system, an uncertain state can be added into the system to express users' uncertainty about service performance. For example, if a rating scheme consists of three states: "good", "fair" and "poor". The UAC assessment scheme, which can be applied in our incentive mechanism, is composed of four states, i.e., "good", "fair", "poor" and "uncertain", where the first three are considered as the *certain* assessments.

7.1.2 Game Setup

In this section, we introduce the broker and payment settings in our work as well users' strategies of giving assessments.

Broker and Payment Settings: the broker requires cloud users to provide continual assessments for services at regular time intervals. A user can get paid by providing an assessment in a scheduled time window. In each time window, only the latest assessment can be paid for by the broker. If the user misses a time window, it cannot give assessments until the next time window. In addition, we assume that the cloud users are long-lived, and care about their long-term payoffs of providing assessments.

In each time window, the broker must pay each user no matter what type of an assessment the user gives. The amount of payment has two levels. If a user gives a certain assessment, it would get a payment P regardless of the value of the assessment. Conversely, if a user gives an uncertain assessment, it would get a discounted payment λP for $\lambda \in [0, 1]$. The reason for why a user can get such a discounted payment is that uncertain assessments cannot benefit the broker but the user still tells the truth without giving untruthful or arbitrary assessments which may even make the broker suffer losses by falsely evaluating the performance of cloud services. If a user does not provide any assessment in a time window, an uncertain assessment would be automatically submitted by a user agent instead.

The compulsory payment setting in our work aims to prevent the broker from "false-reporting" [37]. If the broker can afterwards decide whether to pay according to the quality of assessments, it would always have incentives to refuse to pay to users by cheating about the real quality of assessments. Thus, the payment from the broker in our framework can be considered "*ex-ante*" [193] with two amount levels. The compulsory payment and the judgement of certain or uncertain assessments can be supervised by a *third-party authority* (e.g., a payment management center). The authority can keep both levels of payment (for a/an certain or uncertain assessment) before each time window, and then transfers one level of payment to a user according to the certainty of its assessment, and returns the other level of payment to the broker. Therefore, the broker cannot deny that an assessment is certain or uncertain.

User Strategies: based on our framework, the payoff matrix between the broker and a user in a time window can be specified in Table 7.1. We follow the common assumption of incentive mechanisms made in the literature: a user is rational and self-interested, i.e., every user is motivated to maximize its own payoffs. A user would have three strategies of "*cooperation*", "*semi-cooperation*" or "*non-cooperation*". In our framework, cooperation for a user means giving a truthful assessment; semi-cooperation means giving an uncertain assessment; non-cooperation means giving an untruthful or arbitrary assessment (these two situations will be further discussed separately). B is the benefit a truthful assessment can create for the broker in a time window. P is the full payoff a user can obtain by giving a certain assessment. C is the cost of the effort for a user providing a truthful assessment. In the situations of semi-cooperation and non-cooperation, we consider that a user does not have any cost since it does not try to provide a truthful assessment. We follow the common assumption in the literature of incentive mechanisms, i.e., $B > P > C$. Here, we consider that all users are identical in terms of their knowledge and preference, thus B and C are constant, but P is adjustable. Note that, our work can be easily extended to a situation where there are different types of users by setting suitable system parameters for different users. Table 7.1 indicates that a user's dominant strategy is to always behave non-cooperatively, which is not expected by the broker and cause quite negative effects in cloud performance evaluations.

	User		
	Cooperation	Semi-cooperation	Non-cooperation
Broker	$B-P, P-C$	$-\lambda P, \lambda P$	$-P, P$

Table 7.1: Payoff Matrix in a Time Window

7.1.3 The Assessment Scoring Scheme

In order to make a user's dominant strategy cooperation, we propose an assessment scoring scheme to control users' participation in the transactions of selling their assessments.

In our framework, a user has an assessment score to determine if it can sell its assessments to the broker in a time window. At the end of each time window, a new assessment score will be assigned to each user according to its current score and the submitted assessment. An assessment score θ is a positive integer from a nonempty finite set Θ ($\theta \in \Theta = \{0, 1, 2, \cdots, L\}$), where L is the largest score.

At the end of each time window, the broker can judge whether an assessment is truthful or untruthful through some approaches (e.g., majority opinions). Then it reports its judgement for every user to the authority. According to the broker's reports and users' current assessment scores, the authority updates a new score for every user. Note that, the broker would always report the truth about a user's assessments since the payment is ex-ante and the broker cannot lie about the certainty of an assessment in our framework. However, there may exist an error probability α of the broker falsely reporting without intention, e.g., a truthful assessment is reported as an untruthful one, and vice versa. And α should be smaller than the probability of random guessing, i.e., $\alpha \in [0, 0.5]$.

Let $\tau(\theta, b)$ denote the assessment scoring scheme, and the new score of a user at

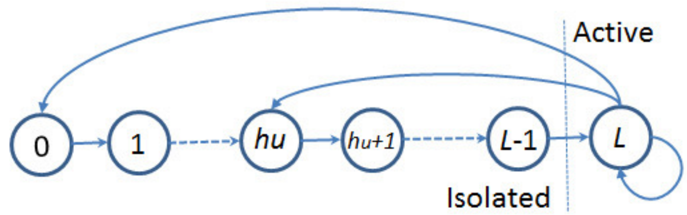


Figure 7.1: The Assessment Scoring Scheme

the end of a time window is computed as follows:

$$\tau(\theta, b) = \begin{cases} L, & \text{if } \theta = L \text{ and } b = T, \\ h_U, & \text{if } \theta = L \text{ and } b = U, \\ 0, & \text{if } \theta = L \text{ and } b = UT, \\ \theta + 1, & \text{if } \theta < L, \end{cases} \tag{7.1}$$

where θ is a user's current score and b is its reported behavior. h_U can be considered as a punishment level for users providing uncertain assessments. A user can be reported as having three types of behaviors, i.e., providing *truthful* (T), *uncertain* (U) or *untruthful* (UT) assessments. Figure 7.1 shows the scoring scheme. If a user having the largest score L is considered to have submitted a/an truthful/uncertain/untruthful assessment, its new score will be maintained at L, or become h_U or 0 respectively, where $0 < h_U < L$. If a user has a score less than L, its score will always increase by 1. Furthermore, the authority requires that only the users having the score L are allowed to submit and sell their assessments to the broker. This means that all users can be classified into two groups: active users and isolated users. If a user is considered to give a/an uncertain or untruthful assessment, it would be punished by being prohibited from selling assessments for a period of time. Thus it will suffer a loss in its future incomes. If a user is not be able to behave cooperatively for some reason, it has a second-best option, i.e., giving uncertain assessments. That is because giving uncertain assessments would cause a shorter period of isolation due to the requirement of $0 < h_U < L$.

7.1.4 Effective Incentive Mechanism Design

In order to build an effective incentive mechanism based on the proposed assessment scoring scheme, we need to analyze the long-term expected payoffs that an "honest" user can obtain and find out what values of L and h_U are necessary for an effective incentive mechanism.

An honest user refers to a user who gives truthful assessments most of the time, but may give a small part of uncertain assessments. We apply the infinite-horizon discounted sum criterion to analyze an honest user's long-term expected payoffs. Let $p(\theta'|\theta)$ denote the transition probability of an honest user's assessment scores between two adjacent time windows, which is shown as follows:

$$p(\theta'|\theta) = \begin{cases} (1-\alpha)(1-\beta), & \text{if } \theta = L \text{ and } \theta' = L, \\ \beta, & \text{if } \theta = L \text{ and } \theta' = h_U, \\ \alpha(1-\beta), & \text{if } \theta = L \text{ and } \theta' = 0, \\ 1, & \text{if } \theta < L \text{ and } \theta' = \theta + 1, \\ 0, & \text{otherwise,} \end{cases} \tag{7.2}$$

where θ represents the user's current score and θ' is the user's new score. α is the error probability of the broker making a false judgement about the user's assessment. β is the probability of the user giving an uncertain assessment in a time window. For an identical type of users and a broker, α and β should be fixed in all time windows. Hence, an honest user's long-term expected payoff in a time window can be computed by solving the following recursive equation:

$$v^{\infty}(\theta) = v(\theta) + \delta \sum_{\theta'} p(\theta'|\theta) v^{\infty}(\theta'), \text{for all } \theta \in \Theta, \tag{7.3}$$

where $v^{\infty}(\theta)$ denotes a user's long-term payoff when it has the assessment score θ in a time window. And $v(\theta)$ denotes the user's instant payoff after giving its assessment in the current time window. $\delta \in (0, 1)$ represents a user's patience about its future pay-

Notations	Explanations
α	The probability for falsely judging an assessment
β	The probability of giving uncertain assessments
γ	The probability for a user guessing correctly
δ	A user's patient for future payoffs
B	The benefit for the broker from a truthful assessment
C	The cost of effort of giving a truthful assessment
P	The ex-ante price for an assessment
λ	The payment discounted factor
L	The largest assessment score
h_U	The assessment score for giving an uncertain assessment

Table 7.2: The Parameters of the Incentive Mechanism

offs. A larger δ means that the user cares more about its future payoffs, and vice versa. Eq. (7.3) indicates that an honest user's long-term expected payoff consists of two parts, i.e., the instant payoff and the expected future payoff based on the score transition probability shown in Eq. (7.2). The notations of our approach are summarized in Table 7.2.

Theorem 1 (Existence of Long-term Expected Payoffs): Given the transition probabilities specified in Eq. (7.2), for any $\alpha \in [0, 0.5]$, $\beta \in [0, 1]$, $\sigma \in (0, 1)$, $\lambda \in [0, 1]$ and $P > C$, the recursive equation Eq. (7.3) has a unique positive solution.

Proof. According to the transition probability in Eq. (7.2), the recursive equation Eq. (7.3) can be expressed as follows:

$$
\begin{aligned}
v^{\infty}(0) &= v(0) + \sigma v^{\infty}(1) \\
&\cdots \\
v^{\infty}(h_U) &= v(h_U) + \sigma v^{\infty}(h_U + 1) \\
&\cdots \\
v^{\infty}(L) &= v(L) + \sigma[(1-\beta)(1-\alpha)v^{\infty}(L) \\
&\quad +(1-\beta)\alpha v^{\infty}(0) + \beta v^{\infty}(h_U)].
\end{aligned}
$$

As the users having assessment scores less than L are isolated, $v(\theta) = 0$ for $\forall\theta < L$.

So we have the following equations:

$$v^\infty(L) = \frac{v^\infty(0)}{\sigma^L},\ v^\infty(h_U) = \frac{v^\infty(0)}{\sigma^{h_U}}.$$

Hence, $v^\infty(0)$ can be computed by solving the following equation:

$$\begin{aligned}\frac{v^\infty(0)}{\sigma^L} &= v(L) + \sigma[(1-\beta)(1-\alpha)\frac{v^\infty(0)}{\sigma^L} \\ &\quad +(1-\beta)\alpha v^\infty(0) + \beta\frac{v^\infty(0)}{\sigma^{h_U}}],\end{aligned}$$

where $v(L) = (1-\beta)(P-C) + \beta\lambda P$, which is the expected instant payoff an honest user can get in a time window. Hence, the solution of Eq. (7.3) is as follows:

$$\begin{aligned} v^\infty(0) &= \frac{\sigma^L(1-\beta)(P-C) + \sigma^L\beta\lambda P}{1-\sigma(1-\beta)(1-\alpha) - \sigma^{L-h_U+1}\beta - \sigma^{L+1}\alpha(1-\beta)}, \\ v^\infty(\theta) &= \frac{v^\infty(0)}{\sigma^\theta},\ \text{for } \forall\theta \in \Theta - \{0\}. \end{aligned}$$

As $\sigma \in (0,1)$, $v^\infty(0) > 0$ according to the conditions in Theorem 1. Thus, Eq. (7.3) has a unique positive solution. □

Based on Theorem 1, we have the following property:

Property 1: The long-term expected payoffs defined in Eq. (7.3) satisfy the following conditions:

(1) $v^\infty(\theta+1) > v^\infty(\theta),$ for $\forall\theta \in \Theta - \{L\}$;

(2) $v^\infty(\theta+1) - v^\infty(\theta) > v^\infty(\theta) - v^\infty(\theta-1),$
for $\forall\theta \in \Theta - \{0, L\}$.

Proof. (1) According to the proof of Theorem 1, $v^\infty(\theta+1) = \frac{v^\infty(0)}{\sigma^{\theta+1}}$ and $v^\infty(\theta) = \frac{v^\infty(0)}{\sigma^\theta}$. As $\sigma \in (0,1)$ and $v^\infty(0) > 0$, the statement (1) is proved.

(2) As $\sigma \in (0, 1)$, $v^{\infty}(0) > 0$ and,

$$
\begin{aligned}
v^{\infty}(\theta+1) - v^{\infty}(\theta) &= \frac{v^{\infty}(0)(1-\sigma)}{\sigma^{\theta+1}}, \\
v^{\infty}(\theta) - v^{\infty}(\theta-1) &= \frac{v^{\infty}(0)(1-\sigma)}{\sigma^{\theta}},
\end{aligned}
$$

the statement (2) is proved. □

In Property 1, the statement (1) indicates that the higher the assessment score of a user, the more the long-term expected payoff. The statement (2) shows that the increase of the long-term expected payoff between two adjacent assessments scores becomes larger with the increase of users' assessment scores. Property 1 demonstrates that an honest user always has incentives to pursue a higher score for obtaining a higher long-term payoff.

In our framework, there should be a dominant strategy for a user, and a second-best strategy if it cannot choose the dominant strategy for some reason. We expect the dominant strategy is to provide truthful assessments, and the second-best strategy is to provide uncertain assessments. As a user's long-term expected payoffs can be computed in a recursive form, its strategy can be determined based on the *one-shot deviation principle* [42], i.e., if a user cannot increase its long-term expected payoff by choosing a strategy other than the dominant one in a time window, the user would not be able to increase the payoff by choosing any strategy other than the dominant one. The one-shot deviation principle can also be applied for the second-best strategy. Hence, we study an active (its assessment score is L) user's long-term expected payoff[4]. If a user provides a *truthful* (T) assessment in a time window, and then its long-term expected payoff can be computed according to Eq. (7.3) as follows:

$$
v_T^{\infty}(L) = P - C + \delta[(1-\alpha)v^{\infty}(L) + \alpha v^{\infty}(0)]. \tag{7.4}
$$

[4]Isolated users are not considered here since such users cannot participate in the transactions of selling assessments until they become active users (their scores increase to L).

And if a user provides an *uncertain* (U) assessment, its payoff can be computed as follows:

$$v_U^\infty(L) = \lambda P + \delta[v^\infty(h_U)]. \tag{7.5}$$

At last, if a user provides an *untruthful* (UT) assessment, its payoff can be computed as follows:

$$v_{UT}^\infty(L) = P + \delta[\alpha v^\infty(L) + (1-\alpha)v^\infty(0)]. \tag{7.6}$$

In order to determine the unique dominant strategy and the second-best strategy, a user's long-term expected payoff should satisfy the constraints: $v_T^\infty(L) > v_U^\infty(L) > v_{UT}^\infty(L)$, i.e.,

$$\begin{aligned} &\delta[(1-\alpha)v^\infty(L) + \alpha v^\infty(0) - v^\infty(h_U)] + (1-\lambda)P - C > 0, \\ &\delta[v^\infty(h_U) - \alpha v^\infty(L) - (1-\alpha)v^\infty(0)] + (\lambda - 1)P > 0. \end{aligned} \tag{7.7}$$

An assessment scoring scheme satisfying Eq. (7.7) indicates that a user can obtain the most long-term expected payoffs when giving a truthful assessment, and the second-best expected payoffs when giving an uncertain assessment.

7.1.4.1 Strategic Users

In Eq. (7.7), we consider that a user only has three kinds of behaviors: providing truthful, uncertain or untruthful assessments. However, there may be *strategic users* who believe that they can guess the real performance of cloud services without actually knowing it. Even for the users who provide *arbitrary* assessments, there should be a small probability that they can guess the right results, so that they would not be punished for "free-riding". The free-riders can be considered as a kind of strategic users. To solve the strategic user problem, we need to reconsider the constraints in Eq. (7.7) for an effective incentive mechanism in our framework.

For strategic users, the computations of the long-term expected payoff of giving a/an truthful or uncertain assessment in a time window are the same as Eqs. (7.4) and (7.5).

Let γ denote the probability that a strategic user (S) guesses the right result of cloud performance. Likewise, the long-term payoff the user can obtain by giving a strategic assessment in a time window is computed as follows:

$$v_S^{\infty}(L) = P + \delta\{\gamma[(1-\alpha)v^{\infty}(L) + \alpha v^{\infty}(0)] + (1-\gamma)[\alpha v^{\infty}(L) + (1-\alpha)v^{\infty}(0)]\}. \quad (7.8)$$

Note that, we only consider the most beneficial case for a strategic user, i.e., a strategic assessment would not incur any cost of effort. Hence, without the consideration of the broker's payoffs, an incentive mechanism is said to be effective if it satisfies all the following constraints:

$$v_T^{\infty}(L) > v_U^{\infty}(L), v_U^{\infty}(L) > v_S^{\infty}(L) \text{ and } v_U^{\infty}(L) > v_{UT}^{\infty}(L). \quad (7.9)$$

Through straightforward calculations, $v_S^{\infty}(L) > v_{UT}^{\infty}(L)$ if and only if $\gamma\alpha < \frac{1}{2}$. In practice, α should usually be in the range of $(0, 0.5)$ (0.5 for random guessing), thus the third constraint in Eq. (7.9) can be omitted in most cases.

7.1.5 Optimal Incentive Mechanism

For a type of users and a broker, there may be many assessment scoring schemes with different parameters L and h_U to satisfy the constraints in Eq. (7.9). In order to find out which parameters are optimal, the total payoffs obtained by both the broker and a user should be analyzed. As only the users having the assessment score L can participate in the transactions of assessments, the total payoffs depend on the proportion of the active users in all users. Let $\eta(\theta)$ denote the proportion of the users having the score θ. Because a user's score is updated at the end of each time window, $\eta(\theta)$ would change dynamically over time. As we assume that users care about their long-term payoffs, we analyze the stationary distribution of $\eta(\theta)$ for $\forall\theta \in \Theta$ if all users are honest. Hence, the stationary distribution can be defined according to the score transition probability in Eq. (7.2) as follows:

$$
\begin{aligned}
\eta(L) &= \eta(L-1) + (1-\alpha)(1-\beta)\eta(L),\\
\eta(\theta) &= \eta(\theta-1), \text{ if } h_U < \theta < L,\\
\eta(h_U) &= \eta(h_U-1) + \beta\eta(L),\\
\eta(\theta) &= \eta(\theta-1), \text{ if } 0 < \theta < h_U,\\
\eta(0) &= \alpha(1-\beta)\eta(L),\\
\sum_{\theta} &\eta(\theta) = 1 \text{ and } \eta(\theta) \geqslant 0, \text{ for } \forall\theta.
\end{aligned}
\tag{7.10}
$$

Theorem 2 (Existence of a Stationary Distribution): Given the transition probabilities specified in Eq. (7.2), for any $\alpha \in [0, 0.5]$, $\beta \in [0, 1]$ and $L > h_U > 0$, there exists a unique stationary distribution satisfying Eq. (7.10).

Proof. According to the definition of stationary distribution in Eq. (7.10), we have:

$$
\begin{aligned}
\eta(0) &= \alpha(1-\beta)\eta(L),\\
&\cdots,\\
\eta(h_U-1) &= \eta(h_U-2),\\
\eta(h_U) &= \eta(h_U-1) + \beta\eta(L),\\
\eta(h_U+1) &= \eta(h_U),\\
&\cdots,\\
\eta(L) &= \eta(L-1) + (1-\alpha)(1-\beta)\eta(L).
\end{aligned}
$$

So we have:

$$
\begin{aligned}
\eta(h_U) &= (\alpha+\beta-\alpha\beta)\eta(L),\\
\eta(h_U) &= \eta(0) + \beta\eta(L),\\
\eta(0) &= \alpha(1-\beta)\eta(L).
\end{aligned}
$$

Since $\sum_\theta \eta(\theta) = 1$, we have:

$$\begin{aligned}
\eta(L) &= \frac{1}{(L-h_U+1)\alpha(1-\beta)+(L-h_U)(\alpha+\beta-\alpha\beta)+1}, \\
\eta(h_U) &= \frac{\alpha+\beta-\alpha\beta}{(L-h_U+1)\alpha(1-\beta)+(L-h_U)(\alpha+\beta-\alpha\beta)+1}, \\
\eta(0) &= \frac{\alpha(1-\beta)}{(L-h_U+1)\alpha(1-\beta)+(L-h_U)(\alpha+\beta-\alpha\beta)+1}.
\end{aligned}$$

According to the conditions specified in Theorem 2, there exists a unique solution for $\eta(L)$, $\eta(h_U)$ and $\eta(0)$. Hence, Theorem 2 is proved. □

Based on Theorem 2, we have the following property:

Property 2: Given the stationary distribution specified in Eq. (7.10), $\eta(L)$ monotonically increases with h_U and monotonically decreases with L.

Proof. According to the unique solution of $\eta(L)$ in the proof of Theorem 2, given α and β, $\eta(L)$ monotonically increases with h_U and monotonically decreases with L if $L > h_U$. □

Property 2 indicates that adjusting L and h_U can change the proportion of active users. The proportion can affect the broker's benefits and users' total benefits.

The expected total payoffs obtained by the broker and an honest user in a time window can be computed as follows:

$$\begin{aligned}
U^* &= \eta(L) \times [(1-\beta)(B-P+P-C)+\beta(-\lambda P+\lambda P)] \\
&= \eta(L) \times (1-\beta)(B-C).
\end{aligned} \tag{7.11}$$

Eq. (7.11) illustrates that U monotonically increases with $\eta(L)$ and decreases with β. In addition, the expected payoff the broker can obtain from an honest user in a time window can be computed as follows:

$$U = \eta(L) \times [(1-\beta)(B-P)-\beta\lambda P]. \tag{7.12}$$

Hence, an effective incentive mechanism in our framework should satisfy the constraints specified in Eq. (7.9) and ensure that the broker can obtain a positive expected payoff in a time window, which is defined as follows:

Definition 1 (Effective Incentive Mechanism): An incentive mechanism with the adjustable parameters L, h_U, λ and P is considered effective if it satisfies the following constraints:

$$\begin{aligned} v_T^\infty(L) > v_U^\infty(L), v_U^\infty(L) > v_S^\infty(L), v_U^\infty(L) > v_{UT}^\infty(L) \\ \text{and } U > 0. \end{aligned} \tag{7.13}$$

Here, we consider maximizing the total payoffs U^* for an optimal incentive mechanism. Thus, we have the following definition:

Definition 2 (Optimal Incentive Mechanism): An effective incentive mechanism is considered optimal if U^* is the maximum for some L, h_U, λ and P.

Note that our work can be simply adjusted for satisfying other targets in any situation, e.g., maximizing the broker's payoff U.

7.1.6 Whitewashing

Whitewashing is a common problem for the reputation or score based incentive mechanisms [41, 192], which refers to the situation where a user can reset its reputation or score by repeatedly re-participating in the activity with new identities. In our scenario, if a user having a score less than L is isolated from assessment transactions, it may try to create a new identity for transactions and expect to come back sooner from the isolation. Here, we assume that a user cannot hold multiple identities at the same time.

By finding out suitable mechanism parameters (i.e., L, h_U and λ), our approach can prevent users from whitewashing. In order to solve this problem, a new user should not enter the assessment transactions instantly. It needs to wait for a period of time as an initializing period, and therefore cannot obtain any benefits. For a new user, an initial assessment score I is assigned. In order to prevent whitewashing, the initial score should satisfy the following constraint:

$$v^{\infty}(I) - v^{\infty}(\theta) \leqslant c_w, \text{ for } \forall \theta \in \Theta \text{ and } I \in \Theta, \tag{7.14}$$

where $c_w \geqslant 0$ is the cost of a user whitewashing, e.g., the cost of creating a new identity. The expression $v^{\infty}(I) - v^{\infty}(\theta)$ indicates the expected long-term gain of a user with the assessment score θ whitewashing. If the gain is no larger than the cost, a user would have no motivation to reset its score. Considering the worst case for preventing whitewashing, i.e., $c_w = 0$, as $v^{\infty}(0)$ is the smallest long-term expected payoff according to the statement (1) of Property 1, $I = 0$ (lowest) is always a solution of Eq. (7.14). Assigning the lowest score to a new user means it can only enter assessment transactions after an initializing period. That means a user with any assessment score cannot gain more payoffs by carrying out whitewashing.

7.2 Illustrative Results and Simulation Results

7.2.1 Parameter Analysis

In our framework, the parameters of an incentive mechanism (see Table 7.2) can be grouped into two classes. The first class includes the intrinsic parameters α, β, γ, δ, B and C. For a type of users and a broker, the intrinsic parameters should be fixed. Thus, an incentive mechanism designer cannot adjust these parameters for an optimal incentive mechanism. The second class includes the adjustable parameters P, λ, L and h_U, where P and λ may need to be conditionally adjusted according to the broker's requirement since they can affect the broker's payoffs.

Fig. 7.2 illustrates the impact caused by α. The vertical axis of the left sub-figure represents the percentage of effective incentive mechanisms in the total number of solutions. Here, we set that L is adjusted from 2 to 10 and λ increases from 0 to 1 by steps of 0.05. The vertical axis of the right sub-figure represents the stationary percentage of active users in the corresponding optimal incentive mechanism. Fig. 7.2

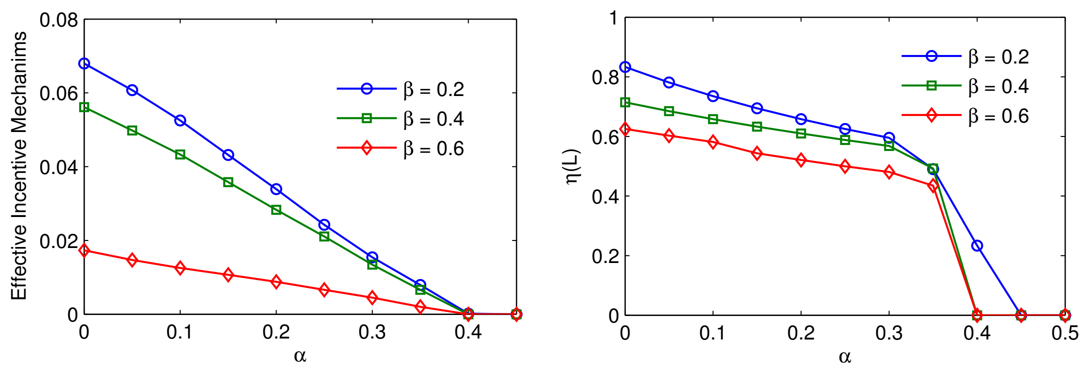


Figure 7.2: Incentive Mechanisms Affected by α

shows that the number of effective incentive mechanisms and active users decrease with α. When α approaches nearly 0.4, there would not be any possible assessment scoring scheme which can be applied to building an effective incentive mechanism, thus the optimal total payoffs (U^*) would be 0. In addition, a larger β would bring a smaller number of active users since an honest user would more often be punished for giving more uncertain assessments. Note that, the maximum possible value of α should only be 0.5 (random guessing) and be much smaller in most of practical cases. In the literature, many approaches are proposed to improve the accuracy of judging assessments for service evaluation, e.g., [101, 117]. Thus, the assumption of the error probability α in our approach is reasonable, so that our work can be applied in most circumstances.

Likewise, Fig. 7.3 shows that the number of effective incentive mechanisms decreases as γ increases. Even if γ reaches a very large value near 0.8, there still exist effective incentive mechanisms, but in those situations, U^* would become very low since the punishment for a strategic user with a high correctness probability should be more serious to prevent its guessing.

Fig. 7.4 demonstrates the results when the price P is adjusted between C and B. When P is near C, the constraints specified in Eq. (7.9) can be hardly satisfied. Conversely, U would be negative when P reaches close to B. Thus, the number of

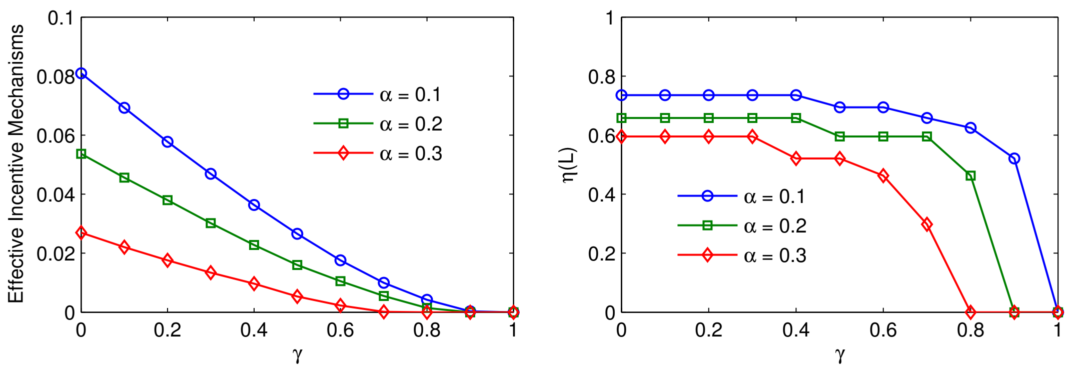


Figure 7.3: Incentive Mechanisms Affected by γ

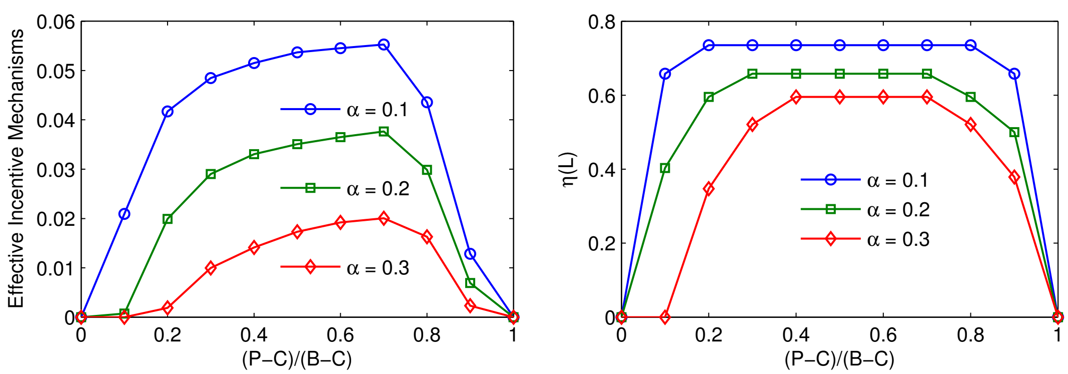


Figure 7.4: Incentive Mechanisms Affected by P

active users would reach the maximum when $\frac{P-C}{B-C}$ is between 0.4 and 0.7 since more effective incentive mechanisms can be built based on such P.

7.2.2 Simulation Experiments

Setting: since there are no suitable real environments supporting our framework, in order to evaluate the real deployment of our work, we have carried out simulation experiments and compared the simulation results with our theoretical results. We have simulated a cloud service environment containing many users, in which a user has its own strategies to provide assessments. Then, we set the same intrinsic parameters for both the simulation environment and the theoretical analysis, and compared the

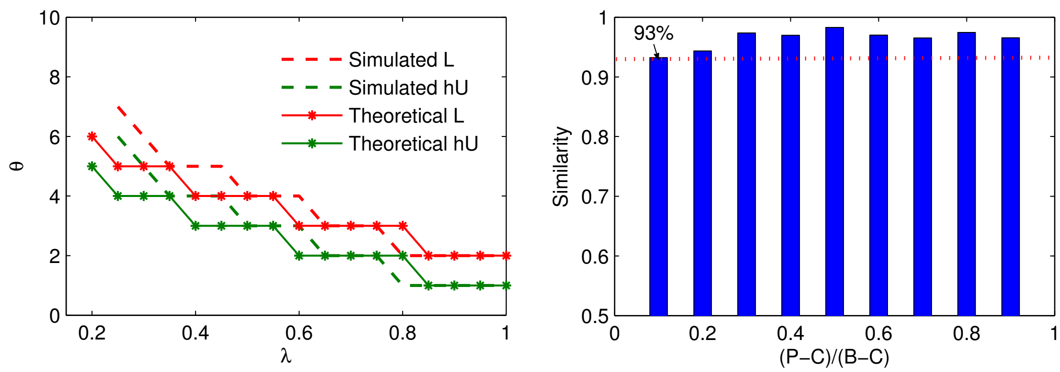


Figure 7.5: Comparison between Simulated Results and Theoretical Results

similarity between the two kinds of results. In the simulation experiments, a setting of the adjustable parameters is considered to build an effective incentive mechanism if, after a number of transactions of selling assessments, a user providing a smaller number of uncertain assessments would get a higher long-term payoff, and a user providing a proportion of uncertain assessments would get a higher long-term payoff than another user providing the same proportion of untruthful or strategic assessments.

Results and Analysis: the left sub-figure of Fig. 7.5 illustrates that the optimal L and h_U between the two kinds of results are very similar when adjusting λ. In some cases, L and h_U in these two kinds of results are not exactly equal since there are unavoidable computational errors in the simulation experiments when taking an action according to a specific probability. If some values of the constraints in Definition 1 are very small but still positive in some assessment scoring schemes, such schemes may be evaluated not to be able to make an effective incentive mechanism in the simulation experiments. Thus, the number of effective incentive mechanisms in the theoretical analysis is usually larger than that in the simulation experiments. According to the experimental results, the average rate between the latter number and the former one is approximately 75%. Likewise, if the values of the constraints are negative but very near 0, such a scheme may be considered to be effective for an incentive mechanism. Even so, the experimental results show that at least 93% of the effective incentive mechanisms in the simulation experiments are the same as those from the theoretical analysis. The right sub-figure of Fig. 7.5 shows such results when P is adjusted between C and B.

7.3 Conclusion

This chapter has proposed a novel incentive mechanism for eliciting continual and truthful assessments in cloud environments. In order to motivate cloud users offering cloud assessments on schedule, this mechanism allows users to provide uncertain assessments in unavoidable situations (e.g., not sure about the real performance of cloud services), and thus protects users' honesty.

Through a suitable assessment scoring scheme, a user would have a dominant strategy (giving truthful assessments) and a second-best strategy (giving uncertain assessments). In the meantime, some optimal goals can be achieved, e.g., maximizing the total payoffs from assessment transactions.

We have theoretically analyzed our approach based on Game Theory and carried out illustrative examples and simulation experiments. The proposed theoretical analysis indicates that our approach is feasible in most circumstances. The simulation experimental results strongly support results from the theoretical analysis.

Chapter 8

Conclusions

Nowadays, cloud computing has become the most popular paradigm for storage and service solutions. The widespread use of cloud computing is also accompanied by some significant issues. One main issue is how to evaluate the performance of cloud services since it is quite necessary for potential cloud consumers to know the quality of services they will consume and pay for. Due to the diversity and dynamic nature of cloud services, selecting the most suitable cloud service has become quite tricky for potential consumers.

This thesis focuses on the two main challenges of cloud service selection. The first one is how to select the most suitable cloud service for potential consumers through comprehensive assessments according to their customized requirements. The other is how to achieve cloud service selection with high effectiveness and accuracy. To this end, four major aspects regarding credible cloud service selection have been studied.

- We have proposed a novel model of cloud service selection by aggregating subjective assessments from ordinary cloud consumers and objective performance assessments from trusted testing parties. In order to consider real world situations, we have applied a fuzzy simple additive weighting system to normalize and aggregate all different types of subjective attributes and objective attributes of cloud services, so that some specific performance aspects of cloud services can also be taken into account according to potential cloud users' requirements. In addition, our model can identify and filter unreasonable subjective assessments. This makes the selection results based on our model more accurate and effective with less noise. The proposed model can effectively reflect cloud consumers' requirements in cloud service selection due to the usage of comprehensive assessments.

- We have proposed a model of context-aware cloud service selection based on comparison and aggregation of subjective assessments and objective assessments. Our model takes the contexts of both subjective assessments and objective assessments into account, and uses objective assessments as a benchmark to filter out unreasonable subjective assessments. The process of such filtering is based on a group of dynamic thresholds which are determined by the similarity between the contexts of subjective assessment and objective assessment. In order to accurately compute the context similarity, we have proposed a novel approach based on the SimRank Algorithm. Our experimental results have shown that our context-aware model performs better than our prior cloud selection model which has no consideration of assessment contexts. Hence, the final aggregated results of cloud services based on our context-aware model can more accurately reflect the overall performance of cloud services.

- In order to reduce the impact caused by biased or noisy assessments, we have proposed a novel model for evaluating cloud users' credibility of providing subjective assessments or objective assessments for cloud services. Our model considers two different classes of cloud users (i.e., ordinary consumers and testing parties). The trustworthiness of ordinary consumers and the reputations of testing parties are computed respectively. And such trustworthiness and reputations can also influence each other, which gives our model the ability to resist user collusion to some extent.

  Inspired by our proposed credibility model, we have proposed *CCCloud*: a credible and context-aware cloud service selection model based on the comparison and aggregation of subjective assessments extracted from ordinary cloud consumers and objective assessments from professional performance monitoring and testing parties. We have first proposed an approach to evaluating the credibility of cloud users. Our approach not only accurately evaluates how truthfully they assess cloud services, but also further resists user collusion. Then, we have considered assessment contexts in cloud service selection. The context similarity between subjective assessments and objective assessments is used to dynamically adjust benchmark levels which are used to filter out biased subjective assessments. The final results of cloud service selection based on assessment credibility and contexts are comprehensive and customized according to the potential consumer's requirements. Hence, our model can satisfy the various needs of different consumers.

We have conducted experiments in many different settings. The experimental results have demonstrated that CCCloud outperforms the existing models under different situations, especially in the situations where the performance of a service is inconsistent, and malicious cloud users give collusive assessments.

- In order to further improve the accuracy of cloud service selection, we have proposed a novel incentive mechanism for eliciting continual and truthful assessments in cloud environments. The main novelty is that, different from prior works, our incentive mechanism is compatible with uncertain assessments. Hence, it can protect a user's honesty by allowing it to give uncertain assessments in unavoidable situations. Through a suitable assessment scoring scheme, a user would have a dominant strategy (giving truthful assessments) and a second-best strategy (giving uncertain assessments). Meanwhile, the total payoffs of transacting assessments would be maximized. We have theoretically analyzed our approach and carried out simulation experiments. The proposed theoretical analysis indicates that our approach is feasible in most circumstances. The simulation experimental results strongly support the theoretical analysis.

Appendix A

The Notations in the Thesis

Table A.1: The Notations in Chapter 3

Notations	Explanations
s	The number of subjective attributes
o	The number of objective attributes
u	The number of associated attributes
$\widetilde{A}$	A trapezoidal fuzzy number
$\mu_{\widetilde{A}}(x)$	The membership function of trapezoidal fuzzy numbers
$d(\widetilde{A})$	The defuzzification method (signed distance)
C_j	The jth alternative cloud service
DM_{jk}	The attribute values of the jth alternative service from the kth decision maker
A_{ijk}	The value of the ith attribute of the jth alternative service from the kth decision maker
$\widetilde{r}_{ijk}$	The fuzzy rating of A_{ijk}
r_{ijk}	The crisp rating of $\widetilde{r}_{ijk}$
ED_{jk}	The Euclidean distance between the ratings of the corresponding subjective associated attributes and the objective associated attributes
$\widetilde{W_i}$	The fuzzy weight for the ith attribute
W_i	The crisp weight for the ith attribute
$\widetilde{M_j}$	The decision matrix for the j alternative service
$\widetilde{S_j}$	The fuzzy score of the jth alternative service
$\overline{S_j}$	The final crisp score of the jth alternative service

Table A.2: The Notations in Chapter 4

Notations	**Explanations**
$s(A, B)$	The similarity between the contexts A and B
$s(c, d)$	The similarity between the assessment features c and d
$V_c(A)$	The value of the assessment feature c in the context A
Cmp_c	The comparator of the assessment feature c
C	The confidence level in the SimRank algorithm
$D(n, n')$	The depth of the deepest common ancestor of two nodes n and n'
$\widetilde{W}_o$	The fuzzy weight for objective assessments
$\widetilde{W}_s$	The fuzzy weight for subjective assessments
g_o	The context of the objective assessments
g_v	The context of the vth group of subjective assessments
R_v	The filtering threshold for the vth group of subjective assessments

Table A.3: The Notations in Chapter 5

Notations	**Explanations**
OC	Ordinary consumers
TP	Testing parties
s_i	The ith alternative cloud service
r	The normalized ratings in the interval $[0, 1]$
$RTr(OC \sim OC')$	The relative trustworthiness of OC' based on OC
R_{TP_q}	TP_q's reputation
$\overline{R_{TP}(OC')}$	The average reputation of similar TPs with OC'
$S_{pri}(OC \sim OC')$	The private similarity between OC and OC'
$S_{pub}(OC' \sim ALL)$	The public similarity between OC' and all other OCs
ω	The weight for private similarity
N_{s_i}	The total number of correspondent rating pairs for s_i
N_{all}	The total number of correspondent rating pairs for all alternative services
N_p	The number of positive correspondent rating pairs for all alternative service
r_{OC,s_i}	The rating of s_i rated by OC
$\rho()$	The Spearman' rank correlation coefficient [103]
$Tr(OC)$	The global trustworthiness of OC
$\varepsilon_a, \varepsilon_b, \varepsilon_c, \varepsilon_d$	The reputation payoffs

Table A.4: The Notations in Chapter 6

Notations	Explanations
CS, TP, OC	Cloud Service, Testing Party, Ordinary Consumer
$Cr_t(OC_i)$	An ordinary consumer OC_i's credibility in the time window t
$Cr_t(TP_j)$	A testing party TP_j's credibility in the time window t
$F_v()$	The factor based on assessment **v**ariation trends (influencing OCs' credibility)
$F_s()$	The factor based on the majority of **s**ubjective assessments (influencing the credibility of OCs and TPs)
$F_o()$	The factor based on the majority of **o**bjective assessments (influencing the credibility of OCs and TPs)
$A_i^s(t)$	OC_i's subjective assessments for all **s**ubjective attributes
$A_i^a(t)$	OC_i's subjective assessments for all subjective **a**ssociated attributes
$M_s^s(t)$	The majority of all OCs' **s**ubjective assessments for all **s**ubjective attributes
$M_s^a(t)$	The majority of all OCs' **s**ubjective assessments for all subjective **a**ssociated attributes
$A_j^o(t)$	TP_j's **o**bjective assessments for all objective attributes
$A_j^a(t)$	TP_j's objective assessments for all objective **a**ssociated attributes
$M_o^o(t)$	The majority of all TPs' **o**bjective assessments for all **o**bjective attributes
$M_o^a(t)$	The majority of all TPs' **o**bjective assessments for all objective **a**ssociated attributes
$F_k(t)$	The final score of CS_k in the time window t

Table A.5: The Notations in Chapter 7

Notations	Explanations
α	The probability for falsely judging an assessment
β	The probability of giving uncertain assessments
γ	The probability for a user guessing correctly
δ	A user's patient for future payoffs
b	A user' reported behavior
B	The benefit for the broker from a truthful assessment
C	The cost of effort of giving a truthful assessment
P	The ex-ante price for an assessment
λ	The payment discounted factor
L	The largest assessment score
h_U	The assessment score for giving an uncertain assessment
$\tau(\theta, b)$	The assessment scoring scheme
$p(\theta\|\theta')$	The transition probability of an honest user's score between time windows
$U_T^\infty(L)$	The long-term expected payoff a user obtains by giving a truthful assessment
$U_U^\infty(L)$	The long-term expected payoff a user obtains by giving an uncertain assessment
$U_{UT}^\infty(L)$	The long-term expected payoff a user obtains by giving an untruthful assessment
$U_S^\infty(L)$	The long-term expected payoff a user obtains by giving a strategic assessment
$\eta(\theta)$	The proportion of the users having the score θ
U^*	The expected total payoffs obtained by the broker and an honest user
U	The expected payoff the broker obtains from an honest user

Appendix B

The Acronyms in the Thesis

Table B.1: The Acronyms in All the Sections

Sections	Explanations	Acronyms
Section 1&5&6	Ordinary Consumer	OC
Section 1&5&6	Testing Party	TP
Section 2	Service-oriented Architecture	SOA
Section 2	Infrastructure-as-a-Service	IaaS
Section 2	Platform-as-a-Service	PaaS
Section 2	Software-as-a-Service	SaaS
Section 2	Service Measurement Indicator	SMI
Section 2	Key Performance Indicator	KPI
Section 3	Fuzzy Simple Additive Weighting System	FSAWS
Section 4	Bias Level	BL
Section 4	Biased Rating Percentage	BRP
Section 6	Context-aware and Credible Cloud Service Selection	CCCloud
Section 6	Cloud Service	CS

Bibliography

[1] Amazon EC2. aws.amazon.com/ec2.

[2] Cloud computing : Tips for financial industry. https://dcvizcayno.wordpress.com/2012/04/13/cloud-computing-tips-for-financial-industry/.

[3] Google Cloud Platform. cloud.google.com/appengine/docs.

[4] Google Docs. www.google.com.au/docs/about.

[5] Microsoft Azure. azure.microsoft.com.

[6] Rackspace. www.rackspace.com.

[7] R. Achar and P. S. Thilagam. A broker based approach for cloud provider selection. In *2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI*, pages 1252–1257, 2014.

[8] M. Alhamad, T. S. Dillon, and E. Chang. SLA-based trust model for cloud computing. In *The 13th International Conference on Network-Based Information Systems*, pages 321–324, 2010.

[9] M. Alrifai, T. Risse, P. Dolog, and W. Nejdl. A scalable approach for QoS-based web service selection. In *Service-Oriented Computing - ICSOC 2008 Workshops, ICSOC 2008 International Workshops, Sydney, Australia, December 1st, 2008, Revised Selected Papers*, pages 190–199, 2008.

[10] M. Alrifai, D. Skoutas, and T. Risse. Selecting skyline services for QoS-based web service composition. In *Proceedings of the 19th International Conference on World Wide Web, WWW*, pages 11–20, 2010.

[11] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. A view of cloud computing. *Commun. ACM*, 53(4):50–58, 2010.

[12] B. Awerbuch, B. Patt-Shamir, D. Peleg, and M. R. Tuttle. Collaboration of untrusting peers with changing interests. In *Proceedings 5th ACM Conference on Electronic Commerce (EC-2004)*, pages 112–119, 2004.

[13] F. Berman, G. Fox, and A. J. G. Hey. *Grid Computing: Making the Global Infrastructure a Reality*. John Wiley & Sons, Inc., 2003.

[14] C. Binnig, D. Kossmann, T. Kraska, and S. Loesing. How is the weather tomorrow?: towards a benchmark for the cloud. In *Proceedings of the 2nd International Workshop on Testing Database Systems, DBTest*, 2009.

[15] R. B. Bohn, J. Messina, F. Liu, J. Tong, and J. Mao. NIST cloud computing reference architecture. In *World Congress on Services, SERVICES 2011, Washington, DC, USA, July 4-9, 2011*, pages 594–596, 2011.

[16] P. A. Bonatti and P. Festa. On optimal service selection. In *International World Wide Web Conferences*, pages 530–538, 2005.

[17] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1-7):107–117, 1998.

[18] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose, and R. Buyya. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Softw., Pract. Exper.*, 41(1):23–50, 2011.

[19] A. R. Callaghan and K. E. Lewis. A 2-phase aspiration-level and utility theory approach to large scale design. In *Proceedings of Design Engineering Technical Conferences And Computers and Information in Engineering Conference*, 2000.

[20] C. Chang, P. Liu, and J. Wu. Probability-based cloud storage providers selection algorithms with maximum availability. In *41st International Conference on Parallel Processing, ICPP 2012, Pittsburgh, PA, USA, September 10-13, 2012*, pages 199–208, 2012.

[21] G. Chen, X. Bai, X. Huang, M. Li, and L. Zhou. Evaluating services on the cloud using ontology qos model. In *IEEE 6th International Symposium on Service Oriented System Engineering, SOSE 2011, Irvine, CA, USA, December 12-14, 2011*, pages 312–317, 2011.

[22] H. Chen, C. Lin, and J. Chen. A multi-objective evolutionary approach for cloud service provider selection problems with dynamic demands. In *Applications of Evolutionary Computation - 17th European Conference, EvoApplications*, pages 841–852, 2014.

[23] Y. Chen, X. Xie, and J. Wu. A benchmark evaluation of enterprise cloud infrastructure. In *Web Technologies and Applications, APWeb*, pages 832–840, 2015.

[24] M. B. Chhetri, S. Chichin, Q. B. Vo, and R. Kowalczyk. Smart cloudbench - automated performance benchmarking of the cloud. In *2013 IEEE Sixth International Conference on Cloud Computing*, pages 414–421, 2013.

[25] C. Choi and H. Jeong. Quality evaluation and best service choice for cloud computing based on user preference and weights of attributes using the analytic network process. *Electronic Commerce Research*, 14(3):245–270, 2014.

[26] S. Chou, Y. Chang, and C. Shen. A fuzzy simple additive weighting system under group decision-making for facility location selection with objective/subjective attributes. *European Journal of Operational Research*, 189(1):132–145, 2008.

[27] B. F. Cooper, A. Silberstein, E. Tam, R. Ramakrishnan, and R. Sears. Benchmarking cloud serving systems with YCSB. In *Proceedings of the 1st ACM Symposium on Cloud Computing, SoCC 2010*, pages 143–154, 2010.

[28] G. Copil, D. Trihinas, H. L. Truong, D. Moldovan, G. Pallis, S. Dustdar, and M. D. Dikaiakos. ADVISE - A framework for evaluating cloud service elasticity behavior. In *Service-Oriented Computing - 12th International Conference, ICSOC*, pages 275–290, 2014.

[29] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson. *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition, 2001.

[30] P. Costa, J. P. Santos, and M. M. da Silva. Evaluation criteria for cloud services. In *2013 IEEE Sixth International Conference on Cloud Computing*, pages 598–605, 2013.

[31] M. M. Deza and E. Deza. *Encyclopedia of Distances*. Springer, 2009.

[32] S. Ding, S. Yang, Y. Zhang, C. Liang, and C. Xia. Combining QoS prediction and customer satisfaction estimation to solve cloud service trustworthiness evaluation problems. *Knowl.-Based Syst.*, 56:216–225, 2014.

[33] D. DiPalantino and M. Vojnovic. Crowdsourcing and all-pay auctions. In *Proceedings 10th ACM Conference on Electronic Commerce (EC-2009)*, pages 119–128, 2009.

[34] Y. Du, X. Wang, L. Ai, and X. Li. Dynamic selection of services under temporal constraints in cloud computing. In *ICEBE*, pages 252–259, 2012.

[35] T. Erl. *Service-Oriented Architecture: Concepts, Technology, and Design*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2005.

[36] R. Fagin, R. Kumar, and D. Sivakumar. Comparing top k lists. *SIAM J. Discrete Math.*, 17(1):134–160, 2003.

[37] M. Feldman, C. H. Papadimitriou, J. Chuang, and I. Stoica. Free-riding and whitewashing in peer-to-peer systems. *IEEE Journal on Selected Areas in Communications*, 24(5):1010–1019, 2006.

[38] D. Fensel, H. Lausen, A. Polleres, J. d. Bruijn, M. Stollberg, D. Roman, and J. Domingue. *Enabling Semantic Web Services: The Web Service Modeling Ontology*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[39] J. Figueira, S. Greco, and M. Ehrgott. *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer Verlag, Boston, Dordrecht, London, 2005.

[40] E. Folkerts, A. Alexandrov, K. Sachs, A. Iosup, V. Markl, and C. Tosun. Benchmarking in the cloud: What it should, can, and cannot be. In *TPCTC*, pages 173–188, 2012.

[41] E. J. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10:173–199, 2000.

[42] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 1991.

[43] S. Galizia, A. Gugliotta, and J. Domingue. A trust based methodology for web service selection. In *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*, pages 193–200, 2007.

[44] J. Gao, P. Pattabhiraman, X. Bai, and W. Tsai. Saas performance and scalability evaluation in clouds. In *IEEE 6th International Symposium on Service Oriented System Engineering, SOSE*, pages 61–71, 2011.

[45] S. K. Garg, S. Versteeg, and R. Buyya. SMICloud: A framework for comparing and ranking cloud services. In *IEEE 4th International Conference on Utility and Cloud Computing, UCC 2011, Melbourne, Australia, December 5-8, 2011*, pages 210–218, 2011.

[46] S. K. Garg, S. Versteeg, and R. Buyya. A framework for ranking of cloud computing services. *Future Generation Comp. Syst.*, 29(4):1012–1023, 2013.

[47] N. Ghosh, S. K. Ghosh, and S. K. Das. SelCSP: A framework to facilitate selection of cloud service providers. *IEEE T. Cloud Computing*, 3(1):66–79, 2015.

[48] L. Gillam, B. Li, and J. O'Loughlin. Benchmarking cloud performance for service level agreement parameters. *IJCC*, 3(1):3–23, 2014.

[49] M. Godse and S. Mulik. An approach for selecting Software-as-a-Service (SaaS) product. In *IEEE International Conference on Cloud Computing, CLOUD*, pages 155–158, 2009.

[50] J. Goffin. On the non-polynomiality of the relaxation method for systems of linear inequalities. *Math. Program.*, 22(1):93–103, 1982.

[51] J. Golbeck. Trust on the world wide web: A survey. *Foundations and Trends in Web Science*, 1(2):131–197, 2006.

[52] A. Goscinski and M. Brock. Toward dynamic and attribute based publication, discovery and selection for cloud computing. *Future Generation Comp. Syst.*, 26(7):947–970, 2010.

[53] B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider, and U. Sattler. Owl 2: The next step for owl. *Web Semant.*, 6(4):309–322, Nov. 2008.

[54] S. Habib, S. Ries, and M. Muhlhauser. Towards a trust management system for cloud computing. In *IEEE Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 933–939, 2011.

[55] J. E. Haddad, M. Manouvrier, and M. Rukoz. TQoS: Transactional and QoS-Aware Selection Algorithm for Automatic Web Service Composition. *IEEE T. Services Computing*, 3(1):73–85, 2010.

[56] C. Hang and M. P. Singh. Trustworthy service selection and composition. *TAAS*, 6(1):5, 2011.

[57] R. Hans, U. Lampe, and R. Steinmetz. QoS-aware, cost-efficient selection of cloud data centers. In *Proceedings of the 2013 IEEE Sixth International Conference on Cloud Computing*, CLOUD '13, pages 946–947, Washington, DC, USA, 2013. IEEE Computer Society.

[58] R. Hans, D. Steffen, U. Lampe, B. Richerzhagen, and R. Steinmetz. Setting priorities - A heuristic approach for cloud data center selection. In *CLOSER 2015 - Proceedings of the 5th International Conference on Cloud Computing and Services Science*, pages 221–228, 2015.

[59] S. Harnisch and P. Buxmann. Evaluating cloud services using methods of supplier selection. In *Business Information Systems - 16th International Conference, BIS*, pages 1–13, 2013.

[60] Q. He, J. Han, Y. Yang, J. Grundy, and H. Jin. QoS-driven service selection for multi-tenant SaaS. In *2012 IEEE Fifth International Conference on Cloud Computing*, pages 566–573, 2012.

[61] I. Horrocks and P. F. Patel-Schneider. A proposal for an OWL rules language. In *Proceedings of the 13th international conference on World Wide Web, WWW*, pages 723–731, 2004.

[62] Z. Hu, L. Zhu, C. Ardi, E. Katz-Bassett, H. V. Madhyastha, J. S. Heidemann, and M. Yu. The need for end-to-end evaluation of cloud availability. In *Passive and Active Measurement - 15th International Conference, PAM*, pages 119–130, 2014.

[63] A. F. M. Huang, C. Lan, and S. J. H. Yang. An optimal QoS-based web service selection scheme. *Inf. Sci.*, 179(19):3309–3322, 2009.

[64] S. Hwang, E. Lim, C. Lee, and C. Chen. Dynamic web service selection for reliable web service composition. *IEEE T. Services Computing*, 1(2):104–116, 2008.

[65] A. Iosup. IaaS cloud benchmarking: approaches, challenges, and experience. In *Proceedings of the 2013 international workshop on Hot topics in cloud services, HotTopiCS*, pages 1–2, 2013.

[66] M. M. Jalloh, S. Zhu, F. Fang, and J. Huang. On selecting composite network-cloud services: a quality-of-service based approach. In *Proceedings of the 2015 Conference on research in adaptive and convergent systems, RACS*, pages 242–246, 2015.

[67] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, New York, NY, USA, 1st edition, 2010.

[68] G. Jeh and J. Widom. Simrank: a measure of structural-context similarity. `http://dbpubs.stanford.edu/pub/2001-41`, 2001.

[69] G. Jeh and J. Widom. SimRank: a measure of structural-context similarity. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26, 2002, Edmonton, Alberta, Canada*, pages 538–543, 2002.

[70] A. Jøsang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.

[71] R. G. Junior, T. Rolim, A. Sampaio, and N. C. Mendonça. A multi-criteria approach for assessing cloud deployment options based on non-functional requirements. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, pages 1383–1389, 2015.

[72] R. Jurca and B. Faltings. An incentive compatible reputation mechanism. In *The Second International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS*, pages 1026–1027, 2003.

[73] R. Jurca and B. Faltings. CONFESS – an incentive compatible reputation mechanism for the online hotel booking industry. In *2004 IEEE International Conference on E-Commerce Technology (CEC 2004)*, pages 205–212, 2004.

[74] R. Jurca and B. Faltings. Collusion-resistant, incentive-compatible feedback payments. In *Proceedings 8th ACM Conference on Electronic Commerce (EC-2007)*, pages 200–209, 2007.

[75] R. Jurca and B. Faltings. Mechanisms for making crowds truthful. *J. Artif. Intell. Res. (JAIR)*, 34:209–253, 2009.

[76] M. Kandori. Social norms and community enforcement. *Review of Economic Studies*, 59(1), 1992.

[77] D. S. Kang, H. Liu, M. P. Singh, and T. Sun. Adaptive process execution in a service cloud: Service selection and scheduling based on machine learning. In *2013 IEEE 20th International Conference on Web Services*, pages 324–331, 2013.

[78] R. Karim, C. Ding, and A. Miri. End-to-end performance prediction for selecting cloud services solutions. In *2015 IEEE Symposium on Service-Oriented System Engineering, SOSE*, pages 69–77, 2015.

[79] R. Karim, C. C. Ding, and A. Miri. An end-to-end QoS mapping approach for cloud service selection. In *IEEE Ninth World Congress on Services, SERVICES*, pages 341–348, 2013.

[80] R. Karim, C. C. Ding, A. Miri, and X. Liu. End-to-end QoS mapping and aggregation for selecting cloud services. In *2014 International Conference on Collaboration Technologies and Systems, CTS*, pages 515–522, 2014.

[81] R. Krebs, C. Momm, and S. Kounev. Architectural concerns in multi-tenant SaaS applications. In *CLOSER 2012 - Proceedings of the 2nd International Conference on Cloud Computing and Services Science*, pages 426–431, 2012.

[82] H. Lausen, J. de Bruijn, A. Polleres, and D. Fensel. The WSML rule languages for the semantic web. In *W3C Workshop on Rule Languages for Interoperability*, 2005.

[83] J. V. Leeuwen. *Handbook of Theoretical Computer Science: Algorithms and Complexity*. MIT Press, Cambridge, MA, USA, 1990.

[84] A. Lenk, M. Menzel, J. Lipsky, S. Tai, and P. Offermann. What are you paying for? performance benchmarking for infrastructure-as-a-service offerings. In *IEEE International Conference on Cloud Computing, CLOUD*, pages 484–491, 2011.

[85] A. Li, X. Yang, S. Kandula, and M. Zhang. Cloudcmp: comparing public cloud providers. In *Proceedings of the 10th ACM SIGCOMM Internet Measurement Conference, IMC*, pages 1–14, 2010.

[86] D. Li, D. Cheung, X. Shi, and V. Ng. Uncertainty reasoning based on cloud models in controllers. *Computers & Mathematics with Applications*, 35(3):99–123, 1998.

[87] L. Li and Y. Wang. A trust vector approach to service-oriented applications. In *2008 IEEE International Conference on Web Services (ICWS 2008), September 23-26, 2008, Beijing, China*, pages 270–277, 2008.

[88] L. Li and Y. Wang. Subjective trust inference in composite services. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI*, 2010.

[89] W. Li, Y. Zhong, X. Wang, and Y. Cao. Resource virtualization and service selection in cloud logistics. *J. Network and Computer Applications*, 36(6):1696–1704, 2013.

[90] X. Li, J. Wu, and S. Lu. QoS-aware service selection in geographically distributed clouds. In *22nd International Conference on Computer Communication and Networks, ICCCN*, pages 1–5, 2013.

[91] Z. Li, L. O'Brien, R. Cai, and H. Zhang. Towards a taxonomy of performance evaluation of commercial cloud services. In *2012 IEEE Fifth International Conference on Cloud Computing, Honolulu, HI, USA, June 24-29, 2012*, pages 344–351, 2012.

[92] G. Liang. Fuzzy MCDM based on ideal and anti-ideal concepts. *European Journal of Operational Research*, 112(3):682–691, 1999.

[93] G.-S. Liang and M.-J. J. Wang. Fuzzy multi-criteria decision-making method for facility site selection. *International Journal of Production Research*, 29(11):2313–2330, 1991.

[94] H. Liang, Y. Du, and S. Li. An improved genetic algorithm for service selection under temporal constraints in cloud computing. In *Web Information Systems Engineering - WISE*, pages 309–318, 2013.

[95] D. Lin, T. Ishida, Y. Murakami, and M. Tanaka. Qos analysis for service composition by human and web services. *IEICE Transactions*, 97-D(4):762–769, 2014.

[96] Y. Liu, A. H. H. Ngu, and L. Zeng. QoS computation and policing in dynamic web service selection. In *Proceedings of the 13th international conference on World Wide Web - Alternate Track Papers & Posters, WWW*, pages 66–73, 2004.

[97] D. Lopez-Pintado. The spread of free-riding behavior in a social network. *Eastern Economic Journal*, 34(4):464–479, 2008.

[98] Y. Ma and C. Zhang. Quick convergence of genetic algorithm for qos-driven web service selection. *Computer Networks*, 52(5):1093–1104, 2008.

[99] Z. Malik and A. Bouguettaya. Evaluating rater credibility for reputation assessment of web services. In *Web Information Systems Engineering (WISE)*, pages 38–49, 2007.

[100] Z. Malik and A. Bouguettaya. Rater credibility assessment in web services interactions. *World Wide Web*, 12(1):3–25, 2009.

[101] Z. Malik and A. Bouguettaya. RATEWeb: Reputation assessment for trust establishment among web services. *VLDB J.*, 18(4):885–911, 2009.

[102] M. Mao and M. Humphrey. A performance study on the VM startup time in the cloud. In *2012 IEEE Fifth International Conference on Cloud Computing*, pages 423–430, 2012.

[103] J. I. Marden. *Analyzing and Modeling Ranking Data*. Chapman & Hall, 1995.

[104] B. Martens and F. Teuteberg. Decision-making in cloud computing environments: A cost and risk based approach. *Information Systems Frontiers*, 14(4):871–893, 2012.

[105] D. Marudhadevi, V. N. Dhatchayani, and V. S. S. Sriram. A trust evaluation model for cloud computing using service level agreement. *Comput. J.*, 58(10):2225–2232, 2015.

[106] W. A. Mason and D. J. Watts. Financial incentives and the ”performance of crowds”. *SIGKDD Explorations*, 11(2):100–108, 2009.

[107] E. M. Maximilien and M. P. Singh. A framework and ontology for dynamic web services selection. *IEEE Internet Computing*, 8(5):84–93, 2004.

[108] E. M. Maximilien and M. P. Singh. Toward autonomic web services trust and selection. In *Service-Oriented Computing - ICSOC*, pages 212–221, 2004.

[109] P. M. Mell and T. Grance. SP 800-145. The NIST Definition of Cloud Computing. Technical report, Gaithersburg, MD, United States, 2011.

[110] M. Menzel and R. Ranjan. Cloudgenius: decision support for web server cloud migration. In *Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, April 16-20, 2012*, pages 979–988, 2012.

[111] M. Menzel, M. Schönherr, and S. Tai. $(MC^2)^2$: criteria, requirements and a software prototype for cloud infrastructure decisions. *Softw., Pract. Exper.*, 43(11):1283–1297, 2013.

[112] N. Miller, P. Resnick, and R. Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373, 2005.

[113] B. Mu, S. Li, and S. Yuan. QoS-aware cloud service selection based on uncertain user preference. In *Rough Sets and Knowledge Technology - 9th International Conference, RSKT*, pages 589–600, 2014.

[114] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation for e-businesses. In *35th Hawaii International Conference on System Sciences HICSS-35*, page 188, 2002.

[115] L. D. Ngan and R. Kanagasabai. OWL-S based semantic cloud service broker. In *2012 IEEE 19th International Conference on Web Services, Honolulu, HI, USA, June 24-29, 2012*, pages 560–567, 2012.

[116] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.

[117] T. H. Noor and Q. Z. Sheng. Credibility-based trust management for services in cloud environments. In *Int'l. Conf. on Service Oriented Computing (ICSOC)*, pages 328–343, 2011.

[118] T. H. Noor and Q. Z. Sheng. Trust as a service: A framework for trust management in cloud environments. In *Web Information Systems Engineering (WISE)*, pages 314–321, 2011.

[119] T. H. Noor, Q. Z. Sheng, and A. Alfazi. Reputation attacks detection for effective trust assessment among cloud services. In *12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom 2013 / 11th IEEE International Symposium on Parallel and Distributed Processing with Applications, ISPA-13 / 12th IEEE International Conference on Ubiquitous Computing and Communications*, pages 469–476, 2013.

[120] T. H. Noor, Q. Z. Sheng, A. H. H. Ngu, A. Alfazi, and J. Law. Cloud armor: a platform for credibility-based trust management of cloud services. In *22nd ACM International Conference on Information and Knowledge Management, CIKM'13*, pages 2509–2512, 2013.

[121] C. Okoli and S. D. Pawlowski. The delphi method as a research tool: An example, design considerations and applications. *Inf. Manage.*, 42(1):15–29, Dec. 2004.

[122] OpenGroup. Service Oriented Architecture : What Is SOA? www.opengroup.org/soa/source-book/soa/soa.htm.

[123] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, November 1999.

[124] T. G. Papaioannou and G. D. Stamoulis. An incentives' mechanism promoting truthful feedback in peer-to-peer systems. In *5th International Symposium on Cluster Computing and the Grid (CCGrid 2005)*, pages 275–283, 2005.

[125] Z. Pawlak, J. Grzymala-Busse, R. Slowinski, and W. Ziarko. Rough sets. *Commun. ACM*, 38(11):88–95, Nov. 1995.

[126] R. Poli, J. Kennedy, and T. Blackwell. Particle swarm optimization. *Swarm Intelligence*, 1(1):33–57, 2007.

[127] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.

[128] H. Qian and Q. Lv. Proximity-aware cloud selection and virtual machine allocation in IaaS cloud platforms. In *Seventh IEEE International Symposium on Service-Oriented System Engineering, SOSE*, pages 403–408, 2013.

[129] H. Qian, H. Zu, C. Cao, and Q. Wang. CSS: facilitate the cloud service selection in IaaS platforms. In *2013 International Conference on Collaboration Technologies and Systems, CTS*, pages 347–354, 2013.

[130] C. Qu and R. Buyya. A cloud trust evaluation system using hierarchical fuzzy inference system for service selection. In *28th IEEE International Conference on Advanced Information Networking and Applications, AINA*, pages 850–857, 2014.

[131] L. Qu, Y. Wang, and M. A. Orgun. Cloud service selection based on the aggregation of user feedback and quantitative performance assessment. In *2013 IEEE International Conference on Services Computing*, pages 152–159, 2013.

[132] L. Qu, Y. Wang, and M. A. Orgun. In *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '16*, 2016.

[133] L. Qu, Y. Wang, M. A. Orgun, L. Liu, and A. Bouguettaya. Cloud service selection based on contextual subjective assessment and objective assessment. In *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14*, pages 1483–1484, 2014.

[134] L. Qu, Y. Wang, M. A. Orgun, L. Liu, and A. Bouguettaya. Context-aware cloud service selection based on comparison and aggregation of user subjective

assessment and objective performance assessment. In *2014 IEEE International Conference on Web Services, ICWS*, pages 81–88, 2014.

[135] L. Qu, Y. Wang, M. A. Orgun, L. Liu, H. Liu, and A. Bouguettaya. Cccloud: Context-aware and credible cloud service selection based on subjective assessment and objective assessment. *IEEE T. Services Computing*, 8(3):369–383, 2015.

[136] L. Qu, Y. Wang, M. A. Orgun, D. S. Wong, and A. Bouguettaya. Evaluating cloud users' credibility of providing subjective assessment or objective assessment for cloud services. In *Service-Oriented Computing - 12th International Conference, ICSOC*, pages 453–461, 2014.

[137] Z. U. Rehman, O. K. Hussain, S. Parvin, and F. K. Hussain. A framework for user feedback based cloud service monitoring. In *International Conference on Complex, Intelligent and Software Intensive System*, pages 257–262, 2012.

[138] J. Repschlaeger, S. Wind, R. Zarnekow, and K. Turowski. Decision model for selecting a cloud provider: A study of service model decision priorities. In *19th Americas Conference on Information Systems, AMCIS*, 2013.

[139] S. Ries, S. M. Habib, M. Mühlhäuser, and V. Varadharajan. Certainlogic: A logic for modeling trust and uncertainty - (short paper). In *Trust and Trustworthy Computing - 4th International Conference, TRUST*, pages 254–261, 2011.

[140] A. Ruiz-Alvarez and M. Humphrey. An automated approach to cloud storage service selection. In *Proceedings of the 2nd International Workshop on Scientific Cloud Computing*, ScienceCloud, pages 39–48. ACM, 2011.

[141] T. Saaty. *Decision Making with Dependence and Feedback The Analytic Network Process*. RWS Publications, Pittsburgh, 1996.

[142] S. Sahri, R. Moussa, D. D. E. Long, and S. Benbernou. DBaaS-Expert: A recommender for the selection of the right cloud database. In *Foundations of Intelligent Systems - 21st International Symposium, ISMIS*, pages 315–324, 2014.

[143] P. Saripalli and G. Pingali. MADMAC: multiple attribute decision methodology for adoption of clouds. In *IEEE International Conference on Cloud Computing, CLOUD*, pages 316–323, 2011.

[144] T. Segaran, C. Evans, J. Taylor, S. Toby, E. Colin, and T. Jamie. *Programming the Semantic Web*. O'Reilly Media, Inc., 1st edition, 2009.

[145] U. Shivakumar, V. Ravi, and G. R. Gangadharan. Ranking cloud services using fuzzy multi-attribute decision making. In *FUZZ-IEEE 2013, IEEE International Conference on Fuzzy Systems*, 2013.

[146] J. Siegel and J. Perdue. Cloud services measures for global use: The service measurement index (SMI). In *2012 Annual SRII Global Conference, San Jose, CA, USA, July 24-27, 2012*, pages 411–415, 2012.

[147] S. S. Skiena. Who is interested in algorithms and why?: Lessons from the stony brook algorithms repository. *SIGACT News*, 30(3):65–74, Sept. 1999.

[148] A. Srivastava and P. G. Sorenson. Service selection based on customer rating of quality of service attributes. In *IEEE International Conference on Web Services (ICWS)*, pages 1–8, 2010.

[149] L. Sun, H. Dong, F. K. Hussain, O. K. Hussain, J. Ma, and Y. Zhang. A hybrid fuzzy framework for cloud service selection. In *2014 IEEE International Conference on Web Services, ICWS*, pages 313–320, 2014.

[150] L. Sun, H. Dong, F. K. Hussain, O. K. Hussain, J. Ma, and Y. Zhang. Multicriteria decision making with fuzziness and criteria interdependence in cloud service selection. In *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE*, pages 1929–1936, 2014.

[151] S. Sundareswaran, A. C. Squicciarini, and D. Lin. A brokerage-based approach for cloud service selection. In *2012 IEEE Fifth International Conference on Cloud Computing*, pages 558–565, 2012.

[152] M. Tajvidi, R. Ranjan, J. Kolodziej, and L. Wang. Fuzzy cloud service selection framework. In *3rd IEEE International Conference on Cloud Networking, CloudNet*, pages 443–448, 2014.

[153] C. Tang and J. Liu. Selecting a trusted cloud service provider for your SaaS program. *Computers & Security*, 50:60–73, 2015.

[154] M. Tang, T. Zhang, J. Liu, and J. Chen. Cloud service QoS prediction via exploiting collaborative filtering and location-based data smoothing. *Concurrency and Computation: Practice and Experience*, 27(18):5826–5839, 2015.

[155] M. Tavakolifard, S. J. Knapskog, and P. Herrmann. Trust transferability among similar contexts. In *Q2SWinet'08 - Proceedings of the 4th ACM Workshop on Q2S and Security for Wireless and Mobile Networks, Vancouver*, pages 91–97, 2008.

[156] V. X. Tran, H. Tsuji, and R. Masuda. A new QoS ontology and its QoS-based ranking algorithm for Web services. *Simulation Modelling Practice and Theory*, 17(8):1378–1398, 2009.

[157] E. Triantaphyllou. *Multi-criteria Decision Making Methods: A Comparative Study*. Springer US, 2000.

[158] D. T. Tsesmetzis, I. Roussaki, and E. D. Sykas. QoS-aware service evaluation and selection. *European Journal of Operational Research*, 191(3):1101–1112, 2008.

[159] Z. ur Rehman, F. K. Hussain, and O. K. Hussain. Towards multi-criteria cloud service selection. In *Proceedings of the Fifth International Conference on In-*

novative Mobile and Internet Services in Ubiquitous Computing, IMIS, pages 44–48, 2011.

[160] Z. ur Rehman, O. K. Hussain, and F. K. Hussain. IaaS cloud selection using MCDM methods. In *ICEBE*, pages 246–251, 2012.

[161] Z. ur Rehman, O. K. Hussain, and F. K. Hussain. Multi-criteria IaaS service selection based on QoS history. In *IEEE International Conference on Advanced Information Networking and Applications*, pages 1129–1135, 2013.

[162] Z. ur Rehman, O. K. Hussain, and F. K. Hussain. Parallel cloud service selection and ranking based on QoS history. *International Journal of Parallel Programming*, 42(5):820–852, 2014.

[163] L. Vu, M. Hauswirth, and K. Aberer. QoS-based service selection and ranking with trust and reputation management. In *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE, OTM Confederated International Conferences CoopIS, DOA, and ODBASE*, pages 466–483, 2005.

[164] A. E. Walsh, editor. *UDDI, SOAP, and WSDL: The Web Services Specification Reference Book*. Prentice Hall Professional Technical Reference, 2002.

[165] H. Wang, C. Lee, and T. Ho. Combining subjective and objective QoS factors for personalized web service selection. *Expert Syst. Appl.*, 32(2):571–584, 2007.

[166] L. Wang, R. Ranjan, J. Chen, and B. Benatallah. *Cloud Computing: Methodology, Systems, and Applications*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 2011.

[167] L. Wang and Z. Wu. A trustworthiness evaluation framework in cloud computing for service selection. In *IEEE 6th International Conference on Cloud Computing Technology and Science, CloudCom*, pages 101–106, 2014.

[168] S. Wang, Z. Liu, Q. Sun, H. Zou, and F. Yang. Towards an accurate evaluation of quality of cloud service in service-oriented cloud computing. *J. Intelligent Manufacturing*, 25(2):283–291, 2014.

[169] X. Wang, J. Cao, and Y. Xiang. Dynamic cloud service selection using an adaptive learning mechanism in multi-cloud computing. *Journal of Systems and Software*, 100:195–210, 2015.

[170] Y. Wang, Q. He, and Y. Yang. QoS-aware service recommendation for multi-tenant SaaS on the cloud. In *2015 IEEE International Conference on Services Computing, SCC*, pages 178–185, 2015.

[171] Y. Wang, L. Li, and G. Liu. Social context-aware trust inference for trust enhancement in social network based recommendations on service providers. *World Wide Web*, 18(1):159–184, 2015.

[172] Y. Wang and V. Varadharajan. Role-based recommendation and trust evaluation. In *9th IEEE International Conference on E-Commerce Technology (CEC 2007) / 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services EEE*, pages 278–288, 2007.

[173] J. Weng, C. Miao, and A. Goh. Protecting online rating systems from unfair ratings. In *Trust, Privacy and Security in Digital Business: Second International Conference, TrustBus*, pages 50–59, 2005.

[174] P. Wieder, J. M. Butler, W. Theilmann, and R. Yahyapour. *Service Level Agreements for Cloud Computing*. Springer Publishing Company, Incorporated, 2011.

[175] J. Witkowski. Eliciting honest reputation feedback in a markov setting. In *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena*, pages 330–335, 2009.

[176] J. Witkowski and D. C. Parkes. Peer prediction without a common prior. In *ACM Conference on Electronic Commerce, EC '12*, pages 964–981, 2012.

[177] E. Wittern, J. Kuhlenkamp, and M. Menzel. Cloud service selection based on variability modeling. In *Service-Oriented Computing - 10th International Conference, ICSOC*, pages 127–141, 2012.

[178] T. Wu, S. Zhang, X. Wu, and W. Dou. A consumer-oriented service selection method for service-based applications in the cloud. In *16th IEEE International Conference on Computational Science and Engineering, CSE*, pages 838–845, 2013.

[179] P. Xiong and Y. Fan. QoS-aware web service selection by a synthetic weight. In *Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD*, pages 632–637, 2007.

[180] Z. Xu, P. Martin, W. Powley, and F. H. Zulkernine. Reputation-enhanced QoS-based web services discovery. In *2007 IEEE International Conference on Web Services ICWS*, pages 249–256, 2007.

[181] J. Yang, W. Lin, and W. Dou. An adaptive service selection method for cross-cloud service composition. *Concurrency and Computation: Practice and Experience*, 25(18):2435–2454, 2013.

[182] S. J. H. Yang, J. S. F. Hsieh, B. C. W. Lan, and J. Chung. Composition and evaluation of trustworthy web services. *IJWGS*, 2(1):5–24, 2006.

[183] H. Yu and S. Reiff-Marganiec. A method for automated web service selection. In *2008 IEEE Congress on Services, Part I, SERVICES I*, pages 513–520, 2008.

[184] Q. Yu. CloudRec: a framework for personalized service recommendation in the cloud. *Knowl. Inf. Syst.*, 43(2):417–443, 2015.

[185] T. Yu and K. Lin. Service selection algorithms for composing complex services with multiple QoS constraints. In *Service-Oriented Computing - ICSOC*, pages 130–143, 2005.

[186] T. Yu, Y. Zhang, and K. Lin. Efficient algorithms for web services selection with end-to-end QoS constraints. *TWEB*, 1(1), 2007.

[187] H. Zhang, Y. Wang, and X. Zhang. Transaction similarity-based contextual trust evaluation in e-commerce and e-service environments. In *IEEE International Conference on Web Services, ICWS*, pages 500–507, 2011.

[188] J. Zhang and R. Cohen. Design of a mechanism for promoting honesty in e-marketplaces. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, pages 1495–1500, 2007.

[189] J. Zhang, R. Cohen, and K. Larson. A trust-based incentive mechanism for e-marketplaces. In *Trust in Agent Societies, 11th International Workshop, TRUST*, pages 135–161, 2008.

[190] M. Zhang, R. Ranjan, S. Nepal, M. Menzel, and A. Haller. A declarative recommender system for cloud infrastructure services selection. In *Economics of Grids, Clouds, Systems, and Services - 9th International Conference, GECON*, pages 102–113, 2012.

[191] X. Zhang, Z. Wang, and F. Xu. Reliability evaluation of cloud computing systems using hybrid methods. *Intelligent Automation & Soft Computing*, 19(2):165–174, 2013.

[192] Y. Zhang, J. Park, and M. van der Schaar. Rating protocols in online communities. *ACM Trans. Economics and Comput.*, 2(1):4, 2014.

[193] Y. Zhang and M. van der Schaar. Reputation-based incentive protocols in crowdsourcing applications. In *IEEE INFOCOM*, pages 2140–2148, 2012.

[194] Z. Zheng, X. Wu, Y. Zhang, M. R. Lyu, and J. Wang. QoS ranking prediction for cloud services. *IEEE Trans. Parallel Distrib. Syst.*, 24(6):1213–1222, 2013.

[195] Z. Zheng, Y. Zhang, and M. R. Lyu. CloudRank: A QoS-driven component ranking framework for cloud computing. In *29th IEEE Symposium on Reliable Distributed Systems (SRDS*, pages 184–193, 2010.