Journal of Network and Computer Applications 81 (2017) 24–36

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Comprehensive and systematic review of the service composition mechanisms in the cloud environments

Asrin Vakili, Nima Jafari Navimipour*

Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran

ARTICLE INFO

Keywords:
Cloud computing
Service
Composition
Review
Systematic

ABSTRACT

Typically, cloud computing includes the provisioning of dynamically scalable and virtualized resources as services over the Internet. In the cloud environment, based on the user's needs, various types of services can be delivered that often must be composited to meet the user requests. Therefore, service composition is emerging as a universal technology in order to integrate distributed and heterogeneous services to combine and consolidate the cloud services. This idea focuses on the innovation of a new cloud service including previously existing services for cost and time reducing and efficiency improving. However, to the best of our knowledge, despite the importance of this matter in cloud environments, there doesn't exist any comprehensive and systematic research and survey in this field. Therefore, the purpose of this paper is to survey the existing techniques and mechanisms which can be addressed in this domain. Briefly, the contributions of this paper are: (1) providing an overview of the existing challenges in a range of problem domains associated with cloud service composition, (2) providing an anatomy of some important techniques throughout scope of cloud service composition and (3) outlining key areas for the improvement of service composition methods in the future research.

1. Introduction

Nowadays, through the Internet and web services (Pooranian et al., 2014) information, data transferring has been facilitated (Navimipour and Zareie, 2015). Cloud computing as the web and Internet-based computing model encompasses the provisioning of dynamically scalable and virtualized resources as services (Bastia et al., 2015; Milani and Navimipour, 2016a, 2016b). Some of the distinguishing characteristics of cloud computing are fast services configuration, elasticity, and scalability (Zhang et al., 2010). It enables an access to remotely configurable computing resources and on-demand hardware and software services providing for minimizing the human efforts needed by the customer as well as improving the service maintaining cost (Ashouraie and Jafari Navimipour, 2015). A cloud service follows pay-as-you-go fashion meaning that customers are merely charged for the time they spend on the service (Chiregi and Navimipour, 2016). Services in cloud computing can be categorized into application and computing services (Armbrust et al., 2009). Regarding the kind of services provided, a cloud might have the form of infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS) (Milani and Navimipour, 2016a, 2016b; Serrano et al., 2015) and expert as a service (EaaS) (Hazratzadeh and Jafari Navimipour, 2017; Jafari Navimipour et al., 2015; Navimipour, 2015; Navimipour et al., 2015; Navin et al., 2014). It is currently being used to deal with challenging problems in different application domains, including industry, science, and government (Candelaa et al., 2014).

In the cloud environment, many types of services need to be provided depending on the user's needs (Jeong et al., 2015). In many applications, the request can be satisfied by integrating and combining some resources called service composition. Service composition as an NP-hard optimization problem (Canfora et al., 2005; Rao and Su, 2005), refers to a larger service providing by services integrating processes (Armbrust et al., 2010). Service composition is emerging as a universal technology in order to integrate some services over the Internet which are distributed and heterogeneous in order to consolidate business applications throughout organization boundaries (Tout et al., 2015). A cloud architecture permits service (Pooranian et al., 2015) composition to answer the users' complicated requests in order to improve the accessibility and flexibility of provided services (Xie et al., 2014). This idea focuses on the offering the new cloud services from previously existing services for cost and time reducing and efficiency improving (Kurdi et al., 2015). Several tasks are usually included in a service composition process each of which corresponds to a service class including many candidate services with the same

* Corresponding author.
E-mail address: jafari@iaut.ac.ir (N.J. Navimipour).

http://dx.doi.org/10.1016/j.jnca.2017.01.005
Received 20 May 2016; Received in revised form 1 January 2017; Accepted 4 January 2017
Available online 05 January 2017
1084-8045/

Table 1
Abbreviation table.

Abbreviation	Definition
ACO-WSC	Ant Colony Optimized- Web Service Composition
ATS	Applied Technical Services
BPEL4WS	Business Process Execution Language FOR Web Service
COM2	Combinatorial Optimization algorithm for cloud COMposition
CSCOS	Cloud Service Composition Optimal-Selection
CSSICA	Classified Search Space Imperialist Competitive Algorithm
DCS	Dynamic Cloud Service selection
DAML-S	DARPA agent markup language for services
ICA	Imperialist Competitive Algorithm
MAS	Multi-Agents Self-organizing
OWL-S	Web Ontology Language for Services
PROCLUS	PROjected CLUStering
QoS	Quality of Service
RQs	Research Questions
SR-CNP	Semi-Recursive Contract Net Protocol
STOCCSC	Service Time Optimization in Cloud Computing Service Composition
SLA	Service-Level Agreement
SOAP	Simple Object Access Protocol
WPC	Windows Performance Counters
WCF	Windows Communication Foundation

functions and different QoS constraints when using a composite service. However, the usage of service's resources and the request of composite service have a strong variability.

On the other hand, to the best of our knowledge, despite the importance of service composition mechanism in the cloud environments, there is not any systematic survey and review about the service composition mechanisms in cloud computing that realizes the need for researchers to do more work on service composition. Therefore, the purpose of this paper is to survey and review the existing service composition techniques in cloud computing comprehensively and systematically. Briefly, the contributions to this paper are as follow:

- Providing an overview of the existing challenges in the range of problem domains associated with cloud service composition.
- Providing a systematic study and overview of the existing techniques for service composition, service selection and other actions that need to integrate the cloud services.
- Providing an anatomy of various pivotal techniques within the scope of cloud service composition.
- Outlining key areas for improving the service composition methods in future research.

The structure of this paper is organized as follows. After the introduction, backgrounds of cloud service composition are provided in Section 2. The related work is analyzed and reviewed in Section 3. The research terminologies and prepares selection mechanisms are provided in Section 4. The selected service composition mechanisms in three main categories are described in Section 5. The taxonomy and comparison of the reviewed mechanisms are presented in Section 6. Also, Section 7 maps out same open issues. Finally, Section 8 concludes the paper. Moreover, Table 1 shows the commonly used abbreviation in the paper.

2. Backgrounds

Cloud computing is an emerging technology that is increasingly being adopted by large and small enterprises in order to attain the top-line development and progress by deriving value from data whereas, at the same time, decreasing the cost of running their IT (Austel et al., 2015). The cloud is changing the manner of data processing and information sharing. For instance, data can be kept on local workplaces and file servers for fast access but face the challenge of sharing it with a large number of people. But, in cloud environments, data can be put into one or more cloud storage systems in order to facilitate their sharing with different other users (Nelson and Peterson, 2013). However, because of fast growth of the number of cloud services, a large number of candidate compositions that would use different services might be used to respond the same query (Benouaret et al., 2014). Also, many types of research aimed at automatic service composition have been included in the literature which can be divided into two diverse categories (Lécué et al., 2008) including those focusing on functional aspects (Klusch et al., 2005) and approaches focusing on process service aspects (Berardi et al., 2003). Service composition can be subdivided into three main steps (Henni and Atmani, 2012) including creation of the process model specifying control and data flow among the activities; discovery, selection and binding of concrete services to every activity in the process model; and execution of the composite service by a coordinating entity (Sivasubramanian et al., 2009). Several research efforts have been done which aimed at providing platforms and languages for service composition (Kaklanis et al., 2016). These efforts can be classified into three main categories including workflow-based approaches (Rao and Su, 2004); XML-based approaches, such as BPEL4WS (Andrews et al., 2003); and ontology-based approaches, such as OWL-S (Burstein et al., 2004) and DAML-S (Sycara et al., 2003).

On the other hand, existing research in QoS representation can have the following categorization: single values representation, multiple values representation, and standard statistical distributions (Zheng et al., 2016). In many mechanisms, each QoS metric is considered as a single value able to be the max, min, or mean value of a QoS. Single values are utilized in QoS-driven service selection or composition approaches. QoS which is represented as a constant value cannot reflect the quality variation. The uselessness of single value representation of service QoS has been recognized, and for representing service QoS standard statistical distributions have been adopted. As a result, QoS need to be viewed as stochastic and service composition problem can be regarded as a decision problem under uncertainty. The response time and prices are modeled as independent, and beta-distributed random variables. The most likely, most pessimistic, and most optimistic values of a service are used to determine the beta distribution. In particular, the QoS metrics are modeled as different distribution functions. In real cases, the QoS distribution of a service can be in any shape, and a well-known statistical distribution is not capable of reflecting the irregularly shaped distribution precisely. The QoS can be obtained through QoS monitoring in three strategies for depending on where the measurement is taken places:

- *Client-side monitoring:* The measurement of QoS is performed on the client side. QoS metric that depends on user experience, such as response time, can be measured on the client side. Response time is measured recording the time gap between a client receiving and sending out a Simple Object Access Protocol (SOAP) message.
- *Server-side monitoring:* The measurement of QoS is done on the server side. This technique needs access to the server for the actual service implementation, not always applicable in practice. Windows Performance Counters (WPC) of the Windows Communication Foundation (WCF) can be used to perform the server-side QoS monitoring.
- *Third-party monitoring:* The measurement of QoS is performed by a third party. The outputs of the QoS monitoring of a service will be periodically provided by the third party. User-independent QoS (e.g. price) is usually identical for different users while user-dependent QoS (e.g. failure probability, response time, throughput, etc.) can widely vary for various users as a result of the unpredictable Internet connections and the heterogeneous user environments.

Moreover, service composition in cloud environments tries to choose and interconnect offered services by different service providers on the basis of a specified business process. As mentioned, the business