Towards Peer Selection in a Semantically-Enriched Service Execution Framework with QoS Specifications

Jun Shen, Ghassan Beydoun
Of Faculty of Informatics
University of Wollongong
Wollongong, Australia
{jshen, beydoun}@uow.edu.au

Brian Henderson-Sellers, Shuai Yuan
Of Faculty of Engineering and Information Technology,
University of Technology, Sydney
Sydney, Australia
brian@it.uts.edu.au, shyuan@eng.uts.edu.au

Graham Low
Of School of Information Systems, Technology and Management
University of New South Wales
Sydney, Australia
g.low@unsw.edu.au

Abstract—This paper promotes an ontology-based multi agent system (MAS) framework to facilitate Peer-to-Peer (P2P) service selection with multiple service properties. P2P-based service has emerged as an important new field in the distributed computing arena. It focuses on intensive service sharing, innovative applications and compositions, and, in some cases, high performance orientation. However, one of the remaining challenges for the P2P-based service composition process is how to effectively discover and select the most appropriate peers to execute the service applications when considering multiple properties of the requested services. By introducing an ontology, different ontology-based e-service profiles can be proposed to facilitate handling multiple properties and to enhance the service oriented process in order to achieve the total or partial automation of service discovery, selection and composition. In this paper, we present a conceptual framework for peer selection with a preliminary mathematical model and a selection process, so as to enhance the P2P-based service coordination system and its components.

Keywords—semantic Web services; quality of service; WSMO; peer-to-peer.

I. INTRODUCTION

With the increasing popularity and growth of Web services, many researchers have been interested in developing effective e-service or e-business applications based on various existing components for agent-based systems [1]. In a multi agent system (MAS) composed of a heterogeneous collection of agents with distinct knowledge-bases and capabilities, coordination and cooperation between agents facilitate the achievement of global goals that cannot be otherwise achieved by a single agent working in isolation [2]. The unique characteristics of a MAS have rendered most standard systems development methodologies inapplicable, leading to the development of Agent Oriented Software Engineering methodologies [3], [4].

However, along with a soaring number of Web services developed in agent-oriented decentralised environments, it is essential to consider the quality of service (QoS) for agents when running business processes. It is obvious that the dynamics and heterogeneity of distributed services become extremely important to both service requestors and service providers. Nevertheless, most research works presented so far are predominantly syntactic and have not truly incorporated semantic ontology approaches for service description and composition within a realistic business context. The discovery and integration of a new service into an existing infrastructure is yet to be fully automated and currently requires significant human effort. As a result, it is problematic that traditional methodologies cannot effectively and autonomously conduct service discovery and composition in a complex dynamic environment. Moreover, the QoS specifications proposed in the literature (e.g., [5], [6], [7]), are yet to agree on common defining concepts.

A set of non-functional properties in Web Service Modelling Ontology (WSMO) [8] ideally can be used as a discriminating factor to refine P2P-based Web services, so as to provide a more reliable service selection in business workflows. In this paper, we present a scalable WSMO-based conceptual framework to describe QoS and other features of Web services in a P2P-based environment. We also sketch an automatic concomitant semantic Web services selection process to automatically find appropriate Web services that effectively fulfil the requestor’s requirements. Hence, we design an approach to select the most appropriate peers that will foster a better service composition according to semantics of the user’s request.

The rest of this paper is structured as follows. Section II illustrates our P2P architecture approach. Section III presents our QoS model and WSMO integration in our P2P framework and sketches a practical solution for selecting appropriate peers with multiple properties, specified by our service quality conceptual model. Section IV is a discussion of other related work, while Section V concludes the paper.
II. P2P MAS TO COMPOSE SEMANTIC WEB SERVICES

Generally, in a P2P MAS architecture dedicated to sharing resources, the MAS acts as an interface to a set of resources distributed across a network. Every user of a resource has an agent acting on his/her behalf. This user interface agent seeks the resource that its master user requests. In addition, every resource, which a user of the system would like to share in return, would have a resource keeper agent that also belongs to the system and acts as a gate keeper to this local repository of resources that it shares with other peer user interface agents as they broadcast their requests. In this architecture, all agents co-operate fulfilling queries and having access to their repository of resources whenever a query received can be assisted by their local resources. Resources shared can be information (files of data, music, etc.) as specified in systems similar to Klampanos and Jose [9] and Mine et al. [10], or alternatively, they can be services as specified in this paper or in [11].

In our proposed P2P framework, the MAS consisting of all cooperating user interface and resource keeper agents respond to requests by a user (e.g., a service requester, a software developer, a human web user) represented by an interface agent in the P2P network that acts on their behalf. In our description of the system here, our focus is not on the agent oriented analysis of such a system, rather it is on the role of a quality of service ontology and domain ontologies in such a system to maintain the function of such a system. Therefore for the sake of simplicity and without loss of generality, we merge the roles of user interface and resource keeper as well as the potential role of history manager into one agent. This agent aims to fulfill the request, e.g., locates services and responds to queries by other similar agents. The collection of all these agents and agents assisting them in their tasks form a P2P community-based cooperative MAS. For composing services using their semantics, a P2P MAS is shown in Figure 1. An agent (an ellipse in Figure 1) representing a user (a hexagon in Figure 1) has access to a knowledge base containing services/resources that the user is willing to share with other users. Each service/file/resources (a cylinder in Figure 1) is identified by a unique identifier within the P2P network (e.g., Service identifier, HTML, pdf, music or video).

As agents automatically interact on behalf of users seeking services to be composed, communities of interest begin to emerge. These communities may overlap. Providers and users of services may belong to more than one community, for instance a service to ‘open an account’ may belong to the community of banking developers as well as that for insurance developers. As more and more services are composed, agents become more efficient and effective by interacting with the agents in the communities most likely to be able to provide them with service components. The P2P system is responsible for locating sites where candidate services are available, based on the previous requests made. The mediation between service requesters and providers is always done by the system. When an agent makes a service request, a candidate agent provider responds either by providing details about services they can supply, or by refusing the service. When all responses are received, the requesting agent combines and refines the results to compose a list of services that can be composed to fulfill the request. The requester agent can then select services it wants to compose and initiates the composition.

Figure 1. The P2P Multi agent system is the collection of the agent assistants and any supporting specialized agents.

After a successful composition a requester’s knowledge base is updated to now include the received and the composed services. Similarly for all agents involved in processing a service request, their knowledge base is also updated with additional information reflecting the domain and attributes of the requester agent. This information is used in future service requests. That is, as agents interact they develop awareness of the services possessed by their peers and which peers may be interested in the services that they themselves have. Each agent keeps a record of its history of service sharing in order to evaluate the quality of service (QoS) and to use this for future service requests. The collection of this history is in essence a distributed QoS ontology distributed across providers. The QoS ontology will provide assessments of past queries and providers. It is
used to produce short lists of candidate nodes for future queries, by calculating the similarity between the current query and its QoS. In a fully evolved P2P system, agents may use this QoS knowledge about other users’ interests to request/negotiate for information from their peers when they do not know who has services of interest. New providers are constantly added to the history, expanding the user-agent’s contact circle.

The proposed strategy of service sharing is domain independent and can be applied to any domain that can be prescribed by an ontology. With appropriate measure of the quality of services located by agents, the proposed P2P service execution system subsequently allows dynamic composition of Web services in a highly distributed and heterogenous computing environment [11] that is adapted from [12] to highlight how ontologies can be used by taking advantage of semantically driven composition of services as is often advocated, e.g., in [13]. We aim to have the system provide, to both service requestors and service providers, the Quality of Service (QoS) evaluation. The system will identify the service providers’ capability and performance so as to enhance the service composition for service clients over the real distributed service network. Due to the complexity of QoS metrics [1], [7], a well-defined QoS service description does not actually exist. With a P2P architecture, the QoS is gauged by a service client through cooperative interactions with other peers that can potentially provide the service. The scope of using ontology-based profiles in this MAS development is possible since most of the current work focuses on the definition of a QoS ontology, vocabulary or measurements, and, to a lesser extent, on a uniform evaluation of qualities. To provide a concrete measure to assess candidate services located by agents, we propose to exploit the Web Service Modelling Ontology (WSMO) [8] as a complementary conceptual framework to create the QoS ontology to describe various perspectives on Web services, thus facilitating the integration of the services. In Section III, we outline the components of the evaluation function, which can be enacted to a specified domain. Subsequently, we envisage Problem Solving Methods [11] as units of analysis corresponding to shared services that can be dynamically selected by agent communication sessions at runtime to best suit the service or the Quality of Service (QoS) required to match the requested service level agreement (SLA). This selection will be made using a Peer to Peer (P2P) searching mechanism to locate appropriate services from other peer agents. Cooperative communication between agents about their existing services, their past service requests (and who fulfilled them in the past) and their performance will enable service requesters to locate the peer service provider with the most suitable QoS.

III. PEER SELECTION WITH WSMO QoS CONCEPTUAL MODEL

First, WSMO defines four high-level notions that relate to semantic Web services, namely Ontologies, Goals, Mediators and Web services. Ontologies: are viewed as formal and explicit specifications of shared conceptualizations [8]. They define a common agreed-upon terminology by providing concepts and relationships among the set of concepts from a real world domain. Goals: are depictions of the expectations a service requestor may have when seeking for a service based on the following aspects: functionality, approach and quality of service. Mediators: coordinate the heterogeneity problem that occurs between descriptions at different levels [14]: data level - different terminologies, protocol level - different communication behaviour between services, and process level - different business processes. WSMO defines four types of mediators: OO Mediators connect and mediate heterogeneous ontologies, GG Mediators connect Goals, WG Mediators link Web services to Goals, and WW Mediators connect Web services resolving mismatches between them. Web services: are descriptions of services that are requested by service requestors, provided by service providers, and agreed between service providers and requesters.

Non-functional properties are usually utilised to describe non-functional aspects such as the creator and the creation date, and to provide natural-language descriptions, etc. All of the four WSMO elements have their own non-functional properties. In this paper, however, our QoS extension is of the same nature as the notion of non-functional properties in “Web services”. In other words, we mainly introduce descriptors of QoS, such as performance, availability, spatial features of distributed services, etc. The incorporated QoS properties could also be used in parallel with existing non-functional attributes proposed by other WSMO elements.

We develop the non-functional properties in WSMO in order to support adaptive P2P-based service composition. These properties are domain-independent and can be used by agents assuming coordinator roles in our framework at runtime (as described in Section II). In using these properties, an emerging organisation of the peer-agent selection process and distribution of tasks is enacted at runtime. The resultant decentralised architecture is coordinated and self-managed effectively with services being allocated to peer/agent hosts, who are able to communicate with each other according to a real business process agreement or standard workflow definitions. In the rest of this paper, we present a more effective and qualitative representation to enable peers to evaluate candidate composition (in Section III.A) and select most appropriate peers (in Section III.B) for a requested service in a P2P information system.
A. A QoS Mode

The P2P-based service selection problem can be generally formulated as the following: Consider $P$ as a set of composing agents, $P = \{p_1, p_2, \ldots, p_N\}$, where each $p_i$ ($i=1$ to $N$) is an agent that gets involved in the composition of a number of services from the set $S$ covering $M$ atomic services, $S = \{s_1, s_2, \ldots, s_M\}$. Each atomic service ($s_j$) cannot be allocated to multiple peers, so let $x_{ij} = 1$ if atomic service ($s_j$) is allocated to Peer ($p_i$) and $x_{ij} = 0$ otherwise, and a Peer ($p_i$) can be allocated with a set of atomic services: $A_{pi} = \{s_1, s_2, \ldots, s_{m_i}\}$. Moreover, let $Q_{p_i,s_j} = \langle RT, CT, AV, RB \rangle$ denote the QoS features of Peer ($p_i$) for atomic service ($s_j$), and $\langle RT, CT, AV, RB \rangle$ represent ResponseTime, ComputationCost, Availability and Reliability. To optimally perform the service composition, the basic objective is to find a set of appropriate peers that makes response time and computation cost as small as possible, while keeping the availability and reliability as large as possible. Therefore, 4 sub-objectives can be defined as $min(\sum RT)$, $min(\sum CT)$, $max(\prod AV)$ and $max(\prod RB)$:

$$O_1 = min(\sum_{p_i \in P} \sum_{s_j} Q_{p_i,s_j}(RT)), s_j \in A_{p_i} \quad (1)$$

$$O_2 = min(\sum_{p_i \in P} \sum_{s_j} Q_{p_i,s_j}(CT)), s_j \in A_{p_i} \quad (2)$$

$$O_3 = max(\prod_{p_i \in P} \prod_{s_j} Q_{p_i,s_j}(AV)), s_j \in A_{p_i} \quad (3)$$

$$O_4 = max(\prod_{p_i \in P} \prod_{s_j} Q_{p_i,s_j}(RB)), s_j \in A_{p_i} \quad (4)$$

However, in order to consider the four objectives as a whole, it can be set as:

$$F = max\left(\sum_{i=1}^{N} w_i \cdot O_i + w_4 \cdot O_4\right)$$

$$\quad \quad \quad w_1 \cdot O_1 + w_2 \cdot O_2$$

(5)

where $\{w_1, w_2, w_3, w_4\}$ denote the weights for the four QoS properties: $RT$, $CT$, $AV$ and $RB$. Subject to the following constraints: $\sum_{j=1}^{M} x_{ij} = 1, \quad j = 1, 2, \ldots, M$ , $\sum_{j=1}^{M} x_{ij} = m_i, \quad i = 1, 2, \ldots, N$ , $\sum_{j=1}^{M} m_i = M$ these respectively ensure there is no conflict between peers to conduct atomic service allocation, and guarantee that the number of allocated atomic services of a peer are valid.

B. Selecting Peers with WSMO Enriched Non-functional Properties

Based on [7], we define an extensible class QoSProperty that aims to extend nonFunctionalProperties class in WSMO for P2P-based service selection.

Class nonFunctionalProperties other existing properties...
hasQoSProperty type QoSProperty
Class QoSProperty sub-Class nonFunctionalProperties
hasPropertyName type string
hasPropertyValue type [int, float, long, others]
hasPreferredValueType type [low, high]
hasWeight type float

Each QoS Property is generally described by PropertyName and PropertyValue. For the purpose of QoS-based selection, two additional properties are defined: hasPreferredValueType and hasWeight. The hasPreferredValueType property represents the desired trend. For example, the lower the response time is, the better QoS that could be achieved. The hasWeight is a value denoting the weight of the property, especially when synthetically measuring several different property metrics. In this context we define the weight value within the range $[0, 1]$, while different end users may have different weight values for their service requirements.

For instance, a peer’s “ResponseTime” can be described in Web service profiles as following:

```
dc__title http://purl.org/dc/elements/1.1# "Peer 1"
dc__description hasValue "ResponseTime for LoanApprove process by peer 1"
hasPropertyType hasValue _string ("ResponseTime")
hasPropertyValue hasValue _int ("500")
hasPreferredValueType hasValue _string ("low")
hasWeight hasValue _float ("0.8")
endNonFunctionalProperties
```

In order to evaluate different non-functional properties of e-service peers, the important concepts in our modelling are: PreferredValueType, Weight and Unified Value. PreferredValueType has two possible values, “low” and “high”. We utilise them to identify two types non-functional properties. For example, “ResponseTime” usually is expected to be as short as possible when choosing an appropriate peer, so the PreferredValueType of “ResponseTime” is “low”. Likewise, “ComputationCost” also usually relates to “low”, as no-one would prefer to choose a service with an expensive computation. However, “Reliability” and “Availability” often fit into the “high” category, since their values are often expected to be as high as possible. Accordingly, all peers’ various properties can be categorised into the two types. With regard to “Weight”, it indicates the importance and priority of certain properties during the service composition, so that
the weight value varies from service to service and from property to property. Lastly, “Unified Value” gives each peer’s overall quality measure, which can be used to assess each peer’s capability to meet the requirements of a requested service.

To enable the peers’ coordinating agent to intelligently select peers and plan a whole composition process, we sketch a selection process to assign the atomic services to appropriate peers within the service composition (Figure 2) that addresses the allocation method for multiple peer profile specifications and takes into account the above formulated objective.

```
Start

 Initialise peers and services

 Select an atomic service and find all candidate peers

 Evaluate QoS Properties

 Assign the atomic service

 If all atomic services are allocated

 N

 Y

 End
```

Figure 2. Allocating Atomic Services to Peers.

IV. RELATED WORK AND DISCUSSION

In recent years, the Semantic Web has become a hot topic and many researchers have turned their interests towards it. Functionality and non-functional properties are two essential aspects for semantic Web services. Functionality is used to measure whether this Web service meets all the functional requirements of an anticipated Web service, i.e. Web services matchmaking; while non-functional properties are qualified to evaluate the performance of the Web service. This has been viewed as a sufficient means to distinguish functionally similar Web services. In our previous work [15], [16], we presented a first sketch of the approach, although paying special attention to the extraction of the ontological description of services and the design of the selection process with OWL-S. The previous prototype was limited to a single specification, while it only considers “ResponseTime” as the selection criterion, which was not sufficient for effective services composition. Instead, this paper extends the description of non-functional properties via model-driven WSMO specification, and also presents an approach for the coordinator to automatically identify the best peers through unifying qualities and properties. In the rest of this section, we summarise and compare other approaches in this area. They are aimed at the same goal of easing semantic Web services development for business process management systems.

Recently, QoS-aware service selection and composition have attracted considerable attention. Most related works focus on the development of QoS ontology languages and vocabularies, as well as the identification of various QoS metrics and their measurements with respect to semantic e-services. In [17], the authors have provided a QoS ontology as a complement for the DAML-S [18] ontology in order to provide a better QoS metrics model. Lee et al. [19] and Ran [6] emphasized a definition of QoS aspects and metrics, but have not included the extensible aspects in QoS, such as incorporating Geo features which we proposed in [15]. In [6], all of the possible quality requirements were introduced and divided into several categories, such as runtime-related, transaction support related, configuration management and cost related, and security-related QoS. Both of them present their definitions and possible determinants. Unfortunately, they are all too abstract to suit the implementation requirement. So, they did not tend to present a practical methodology for real service selection. In [5] and [7], the authors focused on the creation of QoS ontology models, which proposed QoS ontology frameworks aimed at formally describing arbitrary QoS parameters. Additionally, [20] and [21] attempted to conduct a proper evaluation framework and proposed QoS-based service selection, despite the authors failing to present a fair and effective evaluation algorithm. Furthermore, [22] and [23] also considered the evaluation of Quality of Service in the context of an overlay network or P2P principles. Based on [15], authors have evaluated our QoS solution, which was based on ACO and the ontology-based solution proposed in this paper. The evaluation results have been reported in [24].

V. CONCLUSION

In this paper, we have proposed a P2P-based service selection framework from the angle of an ontology-based P2P MAS. We described the operation of the P2P MAS and formulated the basic problem of service selection with multiple properties, and augmented the WSMO description by involving typical QoS perspectives. We have also designed a practical approach to facilitate peer selection. Our service peer selection model is expected to be enhanced in the near future by focusing on concrete and detailed geographic features for location-based services, and we will improve our framework for P2P-based workflow under more dynamic circumstances more effectively. Through this effort, we will be extending more complicated and useful specifications (e.g., representing realistic geographical knowledge) as well as protocols to enhance the accessibility, reliability and availability of e-services in decentralised information systems.
REFERENCES


