QoS-Aware Service Composition for Service-Based Systems Using Multi-Round Vickery Auction

Puwei Wang\textsuperscript{2,3}, Ying Zhan\textsuperscript{1*}, Tao Liu\textsuperscript{2,3}, Xiaoyong Du\textsuperscript{2,3}

\textsuperscript{1}School of Information, Guizhou University of Finance and Economics, Guiyang City, China, 550025
\textsuperscript{2}School of Information, Renmin University of China, Beijing, China, 100872
\textsuperscript{*} Corresponding Author: zhanycathy@163.com

Abstract—The service-oriented paradigm offers support for engineering service-based systems based on service composition. QoS (Quality of Service)-aware service composition chooses a set of services to collectively construct a service-based system, while satisfying global QoS constraints and budget restriction. The service providers naturally are self-interested and strive to maximize their own utilities. Existing approaches use iterative combinatorial auction to address the problem. However, truthful bidding is not optimal strategy for service providers in these approaches. In this paper, we propose a multi-round Vickrey auction to choose an optimal service provider for each task while satisfying our global QoS constraints and budget restriction, and show there may exist a Bayesian Nash equilibrium, in which the service providers will not choose strategically to stay silent and will truthfully bid. Finally, the experimental results show that our approach outperforms the existing combinatorial auction-based approaches.

I. INTRODUCTION

Service-based systems can be seen as a evolution from traditional component-based systems. They are composed of services and exposed as services for use through standardized protocols. The designers of service-based systems usually have global QoS (Quality of Service) constraints, such as the constraint of response time, and a budget, that is the total amount of money that we are willing to at most spend. The service-oriented paradigm offers support for constructing the service-based systems based on QoS-aware service composition. We consider the service composition process could include two steps: first, we can use a planning technique \cite{1} to identify an abstract process with a set of tasks in a service-based system. Second, we can construct the service-based system by choosing a set of services to perform the tasks while satisfying the global QoS constraints and budget restriction \cite{2}. In this paper, we assume we have identified an abstract process with a set of tasks in the service-based system. Our objective is to choose a set of services for the tasks to achieve an optimization goal, such as the maximization of the utility of the service-based system.

Existing service composition approaches \cite{3} focus on designing reconfigurable services that offer different QoS such that we can select our desired QoS. The reconfigurable services increase the selection space, thereby assisting us in selecting the best possible QoS to meet our QoS constraints. Different QoS usually requires different service prices. Thus, our objective is to assess such QoS space and obtain an optimal solution. We can achieve that objective when we know which service provides the minimum price for the QoS requested. In general, their costs are the minimum acceptable prices. However, the service providers usually are self-interested and would not expose their actual costs, that would significantly factor into our decision making. Thus, we face a decision making problem with incomplete information to construct the service-based systems. Recent works use iterative combinatorial auctions to address this issue. Service providers bid for the tasks, and we determine which are the winning bids. If the global QoS constraints cannot be satisfied based on the current bids, the service providers are allowed to improve their bids. However, in these approaches, truthful bidding is not optimal strategy for the service providers. That may lead to a decrease in the utility of the service-based system.

Contribution This paper makes the following contributions: i) We propose a multi-round Vickrey auction to choose an optimal service provider for each task while satisfying our global QoS constraints and budget restriction. ii) We show there exists a Bayesian Nash equilibrium, in which the service providers will not choose strategically to stay silent and will truthfully bid.

The rest of the paper is organized as follows. Section II introduces related work. Section III gives a problem statement. Section IV presents a multi-round Vickrey auction to QoS-aware service composition. Section V presents the experiment results. Section VI concludes this paper.

II. RELATED WORK

Auction-based approaches are recognized as a promising approach to address the service composition problem with incomplete information \cite{4}, \cite{5}. Shamimabi et al. \cite{6} run a single-item auction to choose a service provider for each task. Megha et al. \cite{7}, \cite{8} propose combinatorial auction-based approaches, where the service providers can bid on bundles of tasks. In order to satisfy the QoS constraint, there may be extra rounds where the service providers are allowed to adjust their bids. Naive combinatorial auction for \( n \) items becomes computationally infeasible, with \( 2^n \) possible bundles to price. In order to reduce the large search space, Qiang et al. \cite{9} proposes an approach to filter out uncompetitive bids using
**III. QoS-Aware Service Composition for Service-Based Systems**

We consider a service-based system contains an abstract process with a set of **tasks** \( T = \{t_1, \ldots, t_n\} \) to be executed by concrete services, and there is a set of candidate **services providers** \( S = \{s_1, \ldots, s_m\} \), each in which can provide services to execute at least a task.

Let \( I \) denote a vector of (non-price) QoS attributes, such as response time, throughput, etc. We describe a QoS as a vector \( q = (v_{1}, \ldots, v_{|I|}) \), where \( v_h \) represents a value of \( h \)-th (non-price) QoS attribute in \( I \). For example, let \( I = (\text{response time}, \text{throughput}) \). A QoS \( q \) = (1.8, 4.0) represents that the response time is 1.8 (seconds) and the throughput is 4.9 (mbps). And then, we describe the global QoS constraints as a vector \( q_R = (v_{R,1}, \ldots, v_{R,|I|}) \), where \( v_{R,h} \) represents the worst case acceptable value of the \( h \)-th (non-price) QoS attribute in \( I \). He et al. [9] classifies QoS attributes into two categories. The first is positive QoS attribute. The evaluation of the positive QoS attribute will increase as its value increases. The worst case acceptable value is the maximum acceptable value. Throughput is a typical positive QoS attribute. The second is negative QoS attribute. The evaluation of the negative QoS attribute will decrease as its value increases. The worst case acceptable value is the minimum acceptable value. Response time is a typical negative QoS attribute. We say that a QoS \( q = (v_{1}, \ldots, v_{|I|}) \) satisfies a global QoS constraints \( q_R = (v_{R,1}, \ldots, v_{R,|I|}) \), denoted by \( q \succeq q_R \), while \( \forall h \in [1, |I|], v_h \geq v_{R,h} \), if \( h \)-th attribute is a positive attribute; \( v_h \leq v_{R,h} \), if \( h \)-th attribute is a negative attribute. For example, suppose that the global QoS constraints are that the response time is less than or equal to 2.5 (seconds) and the throughput is greater than or equal to 3.5 (mbps). This can be specified as \( q_R = (2.5, 3.5) \). There is \((1.8, 4.0) \succ (2.5, 3.5)\).

We can obtain a service-based system by finding a set of services to execute its tasks.

**Definition 1:** (Service-Based System) We define a service-based system as \( G := (T, CS, \text{select}, q_G, \delta_G) \), where:

- \( T \) is a set of tasks;
- \( CS \) is a set of services that the service providers could provide for the tasks;
- \( \text{select} : T \to CS \times QoS \times \mathbb{R}^+ \) is a function to select a service for a task, in which QoS represents the set of all possible QoS that the service providers could provide. \( \text{select}(t_i) = (c_{si}, q_i, \delta_i) \) represents that we select a service provider’s service \( cs_i \in CS \) to execute the task \( t_i \in T \). The service provider will provide the QoS \( q_i = (v_{i,1}, \ldots, v_{i,|I|}) \) and \( \delta_i \) is the price of the service provider.
- \( q_G \) is the aggregated QoS of the service-based system;
- \( \delta_G \) is total price, that is the total amount of money that we will pay for the service-based system.

The total price is the sum of the prices of the selected service providers, i.e., \( \delta_G = \sum_{t_i \in T} \delta_i \). The aggregated QoS depends on the QoS provided for the tasks and the structures of the abstract process, such as sequence, parallel, conditional branch and loop. In this paper, we denote the aggregated QoS by \( q_G = \text{aggregate}(\{q_1, \ldots, q_T\}) \), where \( q_i \) is the QoS that the selected service provider offers for the task \( t_i \). Alrifai et al. [13] proposes a set of aggregation functions to compute the aggregated QoS based on the basic structures. For example, suppose there only exists sequential structure and \( I = (\text{response time}, \text{throughput}) \). The aggregated QoS could be \( q_G = (\sum_{t_i \in [1, |T|]} v_{i,1}, \min_{t_i \in [1, |T|]} v_{i,2}) \).

In order to evaluate the **performance** of a service-based system based on its multiple QoS attributes, we normalize the QoS \( q_i \) that the service providers offer for a task \( t_i \in T \) as a single value, denoted by \( \eta(q_i) \), by applying the simple additive weighting technique [14]. Suppose a service-based system \( G \) contains a set of tasks \( T \). We can evaluate its performance as \( u(G) = \sum_{t_i \in [1, |T|]} \eta(q_i) \), and then compute the utility of the service-based system as \( \frac{u(G)}{\delta_G} \).

Suppose a service-based system contains a set of tasks \( T \). Given the global QoS constraints \( q_R \), the budget \( \delta_G \), QoS-aware service composition is to construct the service-based system with a maximum utility.

**Definition 2:** (QoS-aware Service Composition) QoS-aware service composition is to construct a service-based system \( G = (T, CS, \text{select}, q_G, \delta_G) \) with a maximum utility \( \frac{u(G)}{\delta_G} \), while satisfying the global QoS constraints \( q_R \) and the budget restriction \( \delta_G \). Formally,

\[
\max_{G} \frac{u(G)}{\delta_G}
\]

subject to: \( q_G = \text{aggregate}(\{q_1, \ldots, q_T\}) \succ q_R \)

\[
\delta_G = \sum_{t_i \in T} \delta_i \leq \text{budget}
\]
If we know the precise costs of the service providers, we can construct a service-based system with a maximum utility by making the price of each task equal to the lowest of the service providers’ costs. However, the service providers usually are autonomous and self-interested. They naturally want to get their utilities as high as possible. The utility of a service provider is the difference between the price that it receives and the service cost that it spends. In the case that the service providers expose their actual costs, they will only have zero utility. Every service provider usually will not expose his cost to us and his competitors (other service providers). Thus, we face a problem with incomplete information.

IV. MULTI-Round VICKREY AUCTION

We propose a multi-round Vickrey auction to find a service provider for each task. We first decide a desired QoS and a reserve price (the highest price that we are willing to pay) and then runs a Vickrey auction for each task. In each auction, the service providers decide whether to submit a bid below the reserve price or stay silent. If no service provider bids, we run another Vickrey auction with a new desired QoS and a new reserve price, until there is at least a service provider that bids below the reserve price.

A. Auction Procedure

We consider a service-based system contains a set of tasks $T$. The tasks can be divided into two sets: One is the set of assigned tasks $T_a$. For each assigned task $\forall t_i \in T_a$, we have selected a service provider to execute the task. The other is the set of unassigned tasks $T_u$. For each unassigned task $\forall t_i \in T_u$, no service provider bids yet. There is $T_u \cap T_a = \emptyset$ and $T_u \cup T_a = T$. In the beginning, $T_a = \emptyset$ and $T_u = T$.

In round $k$, for each unassigned task $\forall t_i \in T_u$, we first determine a desired QoS $q_i^k$ and a reserve price $r_i^k$. That means that we demand that a service provider offers at least the desired QoS $q_i^k$ if he decides to bid and the highest price that we will pay is $r_i^k$. Second, by observing the desired QoS and the reserve price $(q_i^k, r_i^k)$, a response strategy of each candidate service provider $\forall s_j \in S$ is to submit a sealed bid $b_{i,j}^k \leq r_i^k$ below the reserve price for the desired QoS $q_i^k$ or stay silent $b_{i,j}^k = null$. Third. If receiving bids, we select the service provider with the lowest bid. If not, we do not select any service provider in this round. For the selected service provider with the lowest bid, the price he receives is the reserve price if no other service providers bid or the price is the second-lowest bid if other service providers also bid. In a word, we run a Vickrey auction with a desired QoS and a reserve price for each task. If we select a service provider to execute the unassigned task $t_i \in T_u$, we move the task from the unassigned set $T_u$ to the assigned set $T_a$.

If there is at least a task for which no service provider bids, i.e., $T_u \neq \emptyset$, we go to round $k+1$. As discussed in [9], we need to take finite rounds to choose service providers for the tasks. Let $K$ be the number of rounds which we can take at most. The process is iterated till $T_u = \emptyset$ or the last round $K$ ends. After the game ends, if we have selected a service provider for each task, we construct a service-based system; Otherwise, we fail to construct the service-based system.

Figure 1 shows a game tree that describes the auction procedure. Suppose a service-based system contains two tasks $t_1$ and $t_2$ and there are two candidate providers $s_1$ and $s_2$.

In round 1, we first determine the desired QoS as $q_1^1 = (1.0, 4.0)$ and $q_2^1 = (3.0, 8.1)$ and the reserve prices as $r_1^1 = 17.5$ and $r_2^1 = 23.6$ for the tasks $t_1$ and $t_2$. As shown in the left branch, the service provider $s_2$ submits a bid as $b_{2,1}^1 = 21.3$ (below the reserve price $r_1^1 = 23.6$) for the desired QoS $q_1^1 = (3.0, 8.1)$ for the task $t_2$. No other service providers bid for the task. Thus, the service provider $s_2$ is selected to execute the task $t_2$ and the service price that it will receive is the reserve price as $r_2^1 = 23.6$. The cost of the service provider $s_2$ is 21.0. The utility of the service provider (the price minus his cost) then will be $23.6 - 21.0 = 2.6$. No service providers bids for the task $t_1$. Thus, the set of unassigned tasks is $T_u = \{t_1, t_2\}$. However, as shown in the right branch, the service provider $s_2$ may choose strategically to stay silent in round 1. He decides to wait for a new desired QoS and a new reserve price for the task $t_2$ in the next round. The unassigned tasks thus still are $T_u = \{t_1, t_2\}$.

In round 2, as shown in the left branch, i.e., case 1), there exists only one unassigned task $t_1$. We determine a new desired QoS as $q_1^2 = (2.0, 4.7)$ and a new reserve price as $r_1^2 = 24.5$. The service provider $s_1$ that submits the lowest bid $b_{1,2}^2 = 21.3$ is selected to execute the task $t_1$ and the service price is the second-lowest bid as $b_{1,1}^2 = 23.9$. Now, we have constructed a service-based system. The total price is $23.6 + 23.9 = 47.5$ and the performance of the service-based system is 1.41. Thus, the utility of the service-based system is $1.41/47.5 = 0.03$. As shown in the right branch, i.e., case 2), there are two unassigned tasks $t_1$ and $t_2$. We determine new desired QoS and reserve prices for these two tasks. And then, the service provider $s_2$, which chooses to stay silent in round 1, submits a bid $b_{2,2}^2 = 17.0$ (below the reserve price $r_2^2 = 24.1$) for the desired QoS $q_2^2 = (4.0, 9.0)$ for the task $t_2$. No other service providers bids for the task. The service provider $s_2$ then is selected to execute the task $t_2$ and the service price is the reserve price as $r_2^2 = 24.1$. His utility then will be 7.1. Similarly, the service provider $s_2$ that submits the lowest bid $b_{1,2}^2 = 21.3$ is selected to execute the task $t_1$ and the service
price is the second-lowest bid as \( b_{1,1}^2 = 23.9 \). We then obtain
a service-based system with the utility as \( 0.86/48.0 = 0.018 \).
We find that the utility of the service provider \( s_2 \) increases from
2.6 to 7.1 by choosing strategically not to bid for the
task \( t_1 \) in round 1. We have to revise the desired QoS and the reserve
rice for the task \( t_2 \) and the utility of the service-based
system then greatly decreases from 0.03 to 0.018. This show
it is necessary to motivate the candidates not to stay silent.

**B. Bayesian Nash Equilibrium of Service Providers**

Our idea is that we estimate the service providers’ costs
of executing a task, and then decide the desired QoS and
reserve price based on the estimated costs. A service provider
has different costs for providing different QoS and different
service providers have different costs for providing a same
QoS. The service cost typically has a relationship with the
QoS and the “types” of service providers \([15]\). We define the
cost-performance index to describe the service provider’s type.

**Definition 3:** (Cost-Performance Index) We define the cost
performance index of a service provider \( s_j \in S \) by \( \theta_{i,j} \in \mathbb{R}^+ \nolimits \) for a task \( t_i \in T \). Higher index indicates lower
cost.

We propose a cost function to estimate a service provider
\( s_j \)'s cost of providing a QoS \( q_i \) for a task \( t_i \) as below.

\[
e_c(\theta_{i,j}, q_i) = \lambda(q_i) \theta_{i,j}
\]

in which \( \lambda(q_i) = a \times \eta(q_i)^c, \eta(q_i) \) is the normalized
value of the QoS \( q_i, a > 0 \) and \( c > 0 \) are coefficients. The coefficients
can be evaluated from past experiences.

In order to maximize the utility of the service-based system,
we want to determine the reserve price for each task based on
the lowest of the service providers’ costs. According to the
cost function (1), we find that a service provider that has the
highest cost-performance index will have the lowest service
cost. We assume that the service providers would report their
true QoS, since some verification approaches \([16]\) are proposed
to verify the QoS that the service providers claim. Given the
highest cost-performance index and the desired QoS, we then
can estimate the lowest service cost based on the cost function.

For a task \( t_i \in T \), in round \( k \), suppose we estimate the
highest of the candidate service providers’ cost-performance
indexes as \( e_i^k > 0 (e_i^k > e_i^{k+1}) \). A service provider may choose
strategically to stay silent. That may leads to a decrease in the
utility. We hope that, in round \( k \), a service provider bids iff
his cost-performance index is greater than or equal to \( e_i^k \). Let
\( K \) be the number of rounds that we can take at most.

**Definition 4:** (Bayesian Nash Equilibrium of Service
Providers) We define a Bayesian Nash equilibrium as: in each
round \( k \in [1, K] \), the response strategy of each candidate service
provider \( s_j \) in \( S \) whose cost-performance index is \( \theta_{i,j} \) is to submit
a bid equal to his cost as \( b_{i,j} = e_i^k \) if \( \theta_{i,j} \geq e_i^k \) or to stay silent if \( \theta_{i,j} < e_i^k \).

In order to compute the Bayesian Nash equilibrium, we
first introduce the following probability functions. We assume
the probability distribution of the service providers’ cost-performance
indexes are public knowledge, which can be
evaluated from past experiences. For each task \( \forall t_i \in T \), we
denote the probability distribution of cost-performance indexes
by \( \Gamma \) over \([0, +\infty) \) with probability density function \( f(x) \).
Suppose there are \( m \) candidate service providers. For a service
provider, the probability that the cost-performance indexes of
other providers are all less than \( \theta \) can be computed as \( P(\theta) = \int_{0}^{\theta} f(x)dx \). The probability that the highest
cost-performance index of other service providers is \( \theta \) can be
computed as \( p(\theta) = \partial P(\theta)/\partial \theta = (m-1) \int_{0}^{\theta} f(x)dx \).

**Lemma 1:** There exists a Bayesian Nash equilibrium of service
providers, if: \( \forall k \in [1, K-1] \)

\[
P(e_i^k)(r_i^k - \frac{\lambda(q_i^k)}{e_i^k}) = P(e_i^{k+1})(r_i^{k+1} - \frac{\lambda(q_i^{k+1})}{e_i^{k+1}})) + \int_{e_i^{k+1}}^{e_i^k} \frac{\lambda(q_i^m)}{x} p(x)dx
\]

\[\eta(q_i^k) \geq \eta(q_i^{k+1}) \]

\[s_i^K = \frac{\lambda(q_i^K)}{e_i^K} \]

**Proof:** Consider a candidate service provider \( s_j \in S \) whose
cost-performance index is \( \theta_{i,j} \).

In round \( k \in [1, K-1] \), we consider the cases: Case 1) \( \theta_{i,j} = e_i^k \). If there is another service provider whose the cost-
performance index is greater than or equal to \( e_i^k \), that service
provider will bid at most \( \frac{\lambda(q_i^k)}{e_i^k} \) and the service provider \( s_j \)
will only have a zero utility. If the cost-performance indexes of
other service providers are all less than \( e_i^k \) (the probability is \( P(e_i^k) \)), those service providers do not bid and the utility of
the service provider \( s_j \) will be \( r_i^k - \frac{\lambda(q_i^k)}{e_i^k} \). Thus, the left-side of Equation (2) represents the expected utility of the service
provider \( s_j \) if it bids in round \( k \). Similarity, the first part of the
right-side of Equation (2) represents the expected utility of the
service provider \( s_j \) if it bids in round \( k+1 \) if the cost-
performance indexes of other service providers are all less
than \( e_i^{k+1} \) (the probability is \( P(e_i^{k+1}) \)). The second part of the
right-side of Equation (2) represents the expected utility of the
service provider \( s_j \) if it bids in round \( k+1 \) if the second highest
cost-performance index is \( x \in (e_i^{k+1}, e_i^k) \) (the probability is \( p(x) \)). In other words, Equation (2) represents that the service
provider will obtain a same expected utility both in round \( k \)
and in round \( k+1 \). Thus, in the case of \( \theta_{i,j} = e_i^k \), the service
provider cannot improve his expected utility by staying silent in
round \( k \) and bidding in round \( k+1 \). We can further infer
that the service provider cannot improve his expected utility
by staying silent and bidding in round \( k+2, k+3,..., K \).

Case 2) \( \theta_{i,j} > e_i^k \). According the cost function (1), given a
desired QoS \( q_i^K \), the cost of the service provider is represented as
\( \frac{\lambda(q_i^K)}{\theta_{i,j}} \). If the service provider \( s_j \) bids in round \( k \),
his expected utility, denoted by \( E(\theta_{i,j}^k) \), is below:

\[
E(\theta_{i,j}^k) = P(e_i^k)(r_i^k - \frac{\lambda(q_i^k)}{\theta_{i,j}}) + \int_{\theta_{i,j}^k}^{\theta_{i,j}^k} \frac{\lambda(q_i^m)}{x} p(x)dx
\]
in which, the first term represents the expected utility if the cost-performance indexes of other service providers are all less than \( e_i^k \) and the second term represents the expected utility if the highest of the cost-performance indexes of other service providers, denoted by \( x \), is greater than \( e_i^k \) but is less than \( \theta_{i,j} \).

Similarly, if the service provider \( s_j \) chooses to stay silent in round \( k \), and there is a next round \( k+1 \). If the service provider bids in round \( k+1 \), his expected utility is below:

\[
\mathbb{E}(\theta_{i,j}^{k+1}) = P(e_i^{k+1})(s_i^{k+1} - \frac{\lambda(q_k^{i+1})}{\theta_{i,j}}) + \int_{x_i^{k+1}}^{\theta_{i,j}} \frac{\lambda(q_k^{i+1})}{x_i^{k+1}} p(x) dx
\]

According to Constraints (2) and (3), there is \( \mathbb{E}(\theta_{i,j}^k) - \mathbb{E}(\theta_{i,j}^{k+1}) > 0 \). In the case of \( e_i^k > \theta_{i,j} \), the service provider \( s_j \) cannot improve his expected utility by staying silent in round \( k \) and bidding in round \( k+1 \). We can further infer that the service provider cannot improve his expected utility by staying silent and bidding in round \( k+2 \), \( k+3 \), ..., \( K \).

Case 3) \( \theta_{i,j} < e_i^k \). If the service provider \( s_j \) bids in round \( k \), his expected utility is below:

\[
\mathbb{E}(\theta_{i,j}^k) = P(e_i^k)(s_i^k - \frac{\lambda(q_k^i)}{\theta_{i,j}})
\]

If the service provider \( s_j \) stays silent in round \( k \) and bids in round \( k+1 \), his expected utility is \( \mathbb{E}(\theta_{i,j}^{k+1}) \):

\[
\begin{align*}
\mathbb{E}(\theta_{i,j}^{k+1}) &= \begin{cases} 
P(e_i^{k+1})(s_i^{k+1} - \frac{\lambda(q_k^{i+1})}{\theta_{i,j}}) + \int_{x_i^{k+1}}^{\theta_{i,j}} \frac{\lambda(q_k^{i+1})}{x_i^{k+1}} p(x) dx, & \text{if } e_i^{k+1} < \theta_{i,j} < e_i^k \\
E(e_i^{k+1})(s_i^{k+1} - \frac{\lambda(q_k^{i+1})}{\theta_{i,j}}), & \text{if } \theta_{i,j} < e_i^{k+1}
\end{cases}
\end{align*}
\]

According to Constraints (2) and (3), there is \( \mathbb{E}(\theta_{i,j}^k) - \mathbb{E}(\theta_{i,j}^{k+1}) < 0 \). In the case of \( \theta_{i,j} < e_i^k \), the service provider \( s_j \) cannot improve his expected utility by bids in round \( k \).

In the last round \( K \), as shown in Equation (4), the reserve price is equal to the cost of the service provider whose cost-performance index is \( e_i^K \). If \( \theta_{i,j} > e_i^K \), the service provider’s cost will be less than or equal to the reserve price \( \delta_i^q = \frac{\lambda(q_k^i)}{e_i^k} \).

He can obtain a non-negative utility if he bids or he has a zero utility if he stays silent. Similarly, if \( \theta_{i,j} < e_i^k \), the service provider can only obtain a negative utility if he bids, or he has a zero utility if he stays silent.

We run a Vickrey auction to choose a service provider for the task in each round. The important property of the Vickrey auction is that each service provider cannot improve his utility by submitting a bid not equal to his cost if it decides to bid. In summary, there exists a Bayesian Nash equilibrium of service providers, if Constraints (2), (3) and (4) hold.

In round \( k \), the expected performance \( \mathbb{E}(u(G)) \) and the expected price \( \mathbb{E}(\delta_G) \) of the service-based system \( G \) can be computed as below:

\[
\mathbb{E}(u(G)) = \sum_{t_i \in T_u} (1 - P(e_i^k)) \eta(q_i^k) + \sum_{t_i \in T_u} \eta(q_i^*)
\]

\[
\mathbb{E}(\delta_G) = \sum_{t_i \in T_u} (1 - P(e_i^k)) \delta_i^* + \sum_{t_i \in T_u} \delta_i^*
\]

in which, \( 1 - P(e_i^k) \) is the probability that the highest of the service providers’ cost-performance indexes is greater than or equal to \( e_i^k \), \( \eta(q_i^k) \) is the normalized value of the QoS \( q_i^k \), and \( \delta_i^* \) is the price of providing the QoS \( q_i^k \). For an assigned task \( t_i \), \( q_i^* \) represents the QoS that the selected service provider will provide and \( \delta_i^* \) represents the price of the selected service provider.

Consider the global QoS constraint \( q_R \) and the budget \( \text{budget} \), in round \( k \), we determine the desired QoS \( q_i^k \), the estimate \( e_i^k \) of the service providers’ highest cost-performance index and the reserve price \( \delta_i^k \) for each unassigned task \( \forall t_i \in T_u \) by solving the following utility maximization problem.

\[
\max_{(q_i^k, e_i^k, \delta_i^k) \mid t_i \in T_u} \frac{\mathbb{E}(u(G))}{\mathbb{E}(\delta_G)}
\]

subject to: Constraints (2), (3) and (4)

\[
q_G = \arg\max((\bigcup_{t_i \in T_u} q_i^k) \cup (\bigcup_{t_i \in T_u} q_i^*) - q_R)
\]

\[
\delta_G = \sum_{t_i \in T_u} \delta_i^* + \sum_{t_i \in T_u} \delta_i^k \leq \text{budget}
\]

in which the constraints (2), (3) and (4) guarantee there exists a Bayesian Nash equilibrium of service providers. The constraints (5) and (6) guarantee the satisfaction of the global QoS constraints \( q_R \) and the budget restriction \( \text{budget} \).

V. Evaluation

To the best of our knowledge, there is no available dataset that describes the costs of service providers. Thus, we generate data of the cost-performance indexes and the costs of candidate service providers. Gorbenko et al. [17] shows the response times of real-world Web services usually fit a particular theoretical distribution, such as the gamma distribution, i.e., the number of the Web services that have extremely high or low response time is much less than the number of the Web services that have ordinary response time. First, we choose a gamma distribution to simulate a distribution of the real-world Web services’ cost-performance indexes. We sample randomly the candidate Web services’ cost-performance indexes from the gamma distribution. Second, we generate the costs of candidate Web services based on their cost-performance indexes.

We have implemented an iterative combinatorial auction-based approach [9] for comparison. In this approach, truthful bidding is not optimal strategy for the service providers. In the experiments, for the sake of brevity we describe the global QoS constraints (i.e., a QoS vector) as a single QoS value by applying the simple additive weighting technique [14]. The higher is the QoS value, the more difficult to satisfy is the global QoS constraint. We conducted two sets of experiments. In the first set of experiments, we set the number of tasks in the service-based system at 10. We change the value of global QoS constraints from 9 to 17 while decreasing the budget from 375 to 175. We use them to simulate different difficulty levels of the global constraints. In the second set of experiments, we increase the number of the tasks in the service-based system from 5 to 35 in steps of 5 while the global constraints are
easy to satisfy. In these two sets of experiments, we set the maximum number of rounds to 10. In each subset of the experiments, 100 instances of experiments were run, and the collected results were averaged.

We compare the effectiveness of our approach with that of the combinatorial auction-based approach by the success rate, which is the percentage of scenarios where a solution could be found; the number of rounds taken to obtain a solution; and the utility of the service-based system. The results presented in Figures 2(a) and 2(b) illustrate that the success rate of our approach is better than that of the combinatorial auction-based approach across all the experiments. As shown in Figures 2(c) and 2(d), the combinatorial auction-based approach requires less rounds than our approach when the global constraints are easy to satisfy. The number of rounds required by our approach increases as the number of tasks increases but that required by the combinatorial auction-based approach does not change greatly. We think that the reason is that the service providers can bid on bundles of tasks. But, as the difficulty level of the global constraints increases, the number of rounds required by the combinatorial auction-based approach quickly increases and is almost equal to that required by our approach. Figures 2(e) and 2(f) show the utility of the service-based system. On average, by using our approach, the utility increases from 0.096 to 0.1231. In our approach, the service providers will not choose strategically stay silent and will submit bids equal to their costs. We consider this as the reason why our approach outperforms the combinatorial auction-based approach.

VI. CONCLUSION

We study the problem of choosing one for each task among a set of candidate service providers to maximize the utility of the service-based system, taking account that the candidate service providers are self-interested. We propose a multi-round Vickrey auction to choose a service provider for each task, and show there is a Bayesian Nash equilibrium, in which candidate service providers will truthfully bid. Our experimental results show that our approach outperforms the existing combinatorial auction-based approach.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under grants 61003084, 61232007, and 61472428.

REFERENCES