A NOVEL SEQUENTIAL AUCTION MECHANISM FOR RESOURCE ALLOCATION

1. Mrs. Rama Devi, Associate Professor, from Santhiram Engineering College, 2.T.Lakshmi Madhu Sai,3.N.Mallika 4.J.Lavanya, and 5. S.Lakshmi Prasann 2,3,4,5IV-

B.Tech CSE Students from SREC.

ABSTRACT: **Cloud computing is mainly focused on resource allocation, which requires studying the interactions between customers and cloud managers. However, the recent growth in customer's demands and the exposure of private cloud providers attract the cloud managers to rent extra resources from the CP's so as to handle their backlogged tasks. For the interactions between customers and cloud managers we adopt the options-based sequential auction (OBSA's) to design cloud resource allocation. as compared to previous work our paper can handle customers with heterogeneous demands, provide truthfulness as the dominant strategy, enjoy a simple winner determination procedure and prevent the delayed entrance issue. regarding the interactions between cloud managers and CP's, we provide two parallel markets for resource gathering. We conduct a comprehensive analysis of the two markets and identify the bidding strategies of the cloud managers.**

Key Words: Auction theory, cloud of clouds networks, sequential auctions, options-based sequential auctions, proxy agent, cloud resource allocation, Hamilton-Jacobi-Bellman equation, dynamic markets.

1.Introduction

Modern society relies crucially on efficient processing of the massive amount of data collected from a variety of sources such as customers' information, wireless sensors, and statistical polls, for which cloud computing is a natural platform. Various cloud-based services are offered by different commercial companies such as Microsoft Azure [3], Google Cloud [26], and Amazon EC2 [19].Many companies are also anticipated to join this profitable market by offering cloud services. The recent growth in the customers' demands has motivated the idea of sharing the resources in cloud networks [14], where cloud owners can temporarily rent spare resources from one another to provide better services to the customers. It is anticipated that in the near future, large companies may dominate the entire cloud computing market by renting cloud resources from smaller or private companies. In that case, one of the most suitable candidates for modeling the corresponding cloud resource allocation may be the auction mechanism due to its simplicity, versatility, and a good match with the request and response paradigm in cloud networks. Recently, Amazon Spot Instances is introduced as a simple auction- based framework for resource allocation, where users can bid for their requested cloud servers[4].

1.1 Related works

Auction theory provides a solid mathematical foundation for resource allocation among a set of resource-seeking customers and a set of resource providers. Hence, there exists a body of literature studying auction-based resource allocation in other contexts such as spectrum sharing in cognitive radio networks[28],[4],[3],[1].In modern cloud networks, cloud servers can be classified into different types according to their hardware and software configurations. Also, a bundle of cloud servers of different types may be required to meet the heterogeneous user demands simultaneously. Hence, earlier frameworks that only consider one type of cloud servers and one type of tasks cannot well capture the reality of the market. On the other hand, cloud servers often switch between busy and idle states repeatedly and customers may join and leave the market at will. To capture this dynamism, it is more desirable to hold sequential auctions instead of a single-round auction. One simple approach is to hold a sequence of single-round auctions over time. However, as mentioned in [19], single-round truthful auctions usually lose the truthfulness property when they are extended to sequential auctions. The truthfulness property ensures that customers cannot get higher rewards by manipulating their true valuations for the goods. This consideration motivates us to go beyond the existing works on single-round auction (e.g., [5],[18],[20],[15]), and to seek truthful sequential auction solutions.

The most related works to ours are [6],[29],[24],[8], [9]. In [6], a novel bidding language is introduced based on categorizing the users into different groups with respect to their characteristics. Users are partitioned into three groups: job-oriented users, resource-aggressive users, and resource aggressive users with time-invariant capacity requirements. A truthful online cloud auction mechanism is introduced on top of this bidding language. However, the original model only considers one type of the cloud servers. The authors have extended their proposed frame- work to the case with multiple types of tasks and servers in [17]. However, the resulting model requires calculating a complex payment function for each arriving task and obtaining the allocation strategy by solving an optimization problem. These issues become a concern when handling real-time resource

allocation in cloud networks with a large task arrival rate.

In order to model the multiple types of cloud servers and customers with heterogeneous demands, current literature has mainly focused on utilizing the combinatorial auctions for cloud resource allocation. Although combinatorial auctions can guarantee some favorable properties (such as truthfulness) in theory, it is well-known that determining the winner and its payment in combinatorial auctions is NP-hard, which renders them impractical in dynamic markets with real-time demands such as cloud networks. Also, these auctions are inherently designed for one-round selling. These issues of combinatorial auctions have promoted further research (e.g.[27],[24])on solving winner determination using simpler approximation methods or extending them to sequential combinatorial auctions. In [27], the au-thors proposed a truthful mechanism for sequential combinatorial auctions. In this framework, besides complicated winner determination and payment identification process, when a user's task requires a bundle of cloud resources for more than one unit of time, the user has to bid in multiple rounds of auctions. This fact makes the framework inapplicable when users require uninterruptible processing of their tasks. In [24], the interactions between customers and cloud providers are modeled as an online combinatorial auction. The model of that work captures multiple types of cloud servers and heterogeneity of customers' demands. Also, it considers a sequential style of auction, in which winner determination is translated into a series of one-round optimization problems. A truthful mechanism of selling is examined; and an approximate algorithm is proposed for one-round optimization. However, similar to [17], the need for solving multiple optimization problems using empirical methods in each round of the auction makes the framework complicated and computationally intensive. All of the afore- mentioned works and most of the contemporary literature focus on modeling the interactions between cloud managers and customers, whereas the resource gathering process for cloud managers is largely ignored. In the pioneering work [9], a general framework for inter cloud networks is presented where the interactions between users and cloud providers is modeled by many-to-many auctions.

Afterward, the interactions between cloud providers are modeled by a coalition game in which the cloud providers borrow re-sources from each other to fulfill their customer's demands. This work is among the first to consider the interactions both between the customers and cloud providers, and among the cloud providers themselves; unfortunately, users with bundle demands were not considered. Furthermore, one of the main challenges in dynamic cloud resource allocation scheme neglected in most of the mentioned works is the delayed entrance problem. This problem arises when a user delays its entrance into the market when he has some side-information about the future dynamics of the market. Assume that in a sequential combinatorial auction, some users become aware that by waiting for some period of time, the cloud resources can be obtained at lower prices.

In summary, existing literature on modeling the inter- actions between the cloud managers and customers has at least one of the following four limitations:

1.Incapability the of handling customers' heterogeneous demands that require a bundle of different types of servers.

2.Missing the truthfulness property.

3.Requiring prohibitive computation for winner and payment determination.

4. Susceptible to the delayed entrance issue. These issues will be addressed in our work. Also, to the best of our knowledge, our work is among the first to leverage auction theory to study the interactions among the public cloud managers and private cloud providers(CPs),which better captures the selfishness of the CPs considering their willingness to resource sharing captured by their offered prices.

1.2 Novelty and Contributions

To address the limitations of the existing works mentioned in subsection 1.1, a novel two-stage auction framework is proposed in this work to capture the interactions among. customers and cloud managers, and (b) cloud managers and CPs. Specifically, we consider cloud of clouds networks (CCNs) consisting of heterogeneous cloud servers and customers with different demands. There exists a CCN manager in charge of handling the resources of each CCN. The CCN managers are interested in renting servers from CPs to en- large their pool of resources so as to attract more customers and better handle their real-time demands. The first stage of the proposed framework is inspired by the options-based sequential auctions (OBSAs) [17] and models the interactions between customers and CCN managers, in which each customer endeavors to obtain his/her demanded resources from a CCN.[1]To the best of our knowledge, we are (among) the first to leverage OBSAs to address the major limitations of existing works on dynamic cloud networks.[2]In addition, we provided the corresponding performance analysis based on a novel Markov chain modeling, which is new to existing studies in the relevant literature. The second stage of the proposed framework describes the interactions between multiple CCN managers and multiple CPs, in which CCN managers compete to obtain resources from CPs. For this stage, we introduce a novel model consisting of two parallel markets for gathering cloud resources: flat-price market and auction-based market, to better capture the selfishness of the private CPs (by incorporating offered prices) as compared to the existing models. We also provide a comprehensive analysis for these markets using Hamilton-Jacobi-Bellman equation and derive the bidding strategy of the CCN man- agers with respect to their inherent characteristics in a stable market setting.

Structure of the paper:
The system model is introduced in Section 2.The interactions between the customers and the CCN managers are modeled in Section 3. Section 4 is devoted to the analysis of OBSAs. The interactions between the CPs and the CCN managers is modeled and analyzed in Section 5. Simulation results are presented in Section 6. Finally, Section 7 concludes the paper and provides some possible future directions.

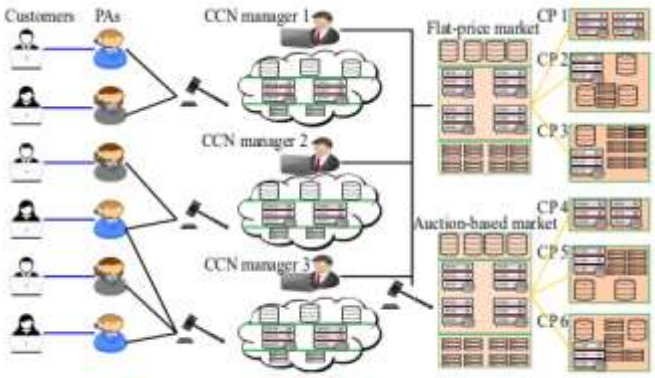

Figure 1: Market model.

2.SYSTEM MODEL

A CCN consists of multiple cloud servers with different processing capabilities; some of them are more desirable for GPU processing, while the others are more suitable for real time database analysis and parallel processing. In addition to their core servers, CCNs can rent servers from CPs to process their backlogged tasks and to serve more customers. CPs are small cloud retailers who lease their extra computational resources to CCNs for profit. Customers with multiple heterogeneous demands may join the CCN at will and require multiple types of servers simultaneously. Inspired by [22], [17], proxy agents (PA) are incorporated into our model as trusted mediators between the customers and the corresponding cloud of cloud networks. Each customer sends its demands to an idle PA; subsequently, the PA attempts to fulfill the demands with the available resources of a CCN. Each CCN operates under the control of a CCN manager who interacts with CPs and PAs.

Due to the variety in the task types and individual priorities, customers may have disparate preferences for different (combination of) servers, which is assumed known to their corresponding PAs. In this paradigm, the PAs and the CCN managers employ a common bidding language that reflects the customers' demands and valuations. Nevertheless, the discussion of the bidding language is beyond the scope of this paper. An interested reader is referred to [10], [24] and references therein for more details.

In this work, we introduce a framework in which CCN managers rent extra servers from CPs by participating in one of the two parallel markets: the flat-price market and the auction-based market. In the flat-price market, the CPs offer their servers at a fixed price. In the auction-based market, CPs provide their servers along with their offered prices (i.e., the least expected price to lend the corresponding servers), where the CCN managers bid in a sequence of auctions to obtain the servers while satisfying the CPs' offered prices. The flat-price market is more suitable for leasing the servers with long idle periods. In this case, since

A CP does not need to utilize the server in the near future, he aims to lease the server with a constant high price. However, the auction-based market is more favorable for servers with a short idle period. In this case, the CP may need its servers in a near future for itself. Hence, CPs compete with each other by offering lower prices of their servers so as to lease them faster.

Similarly, a CCN manager who requires the resource immediately and needs to rent it for a long period tends to join the flat-price market, while the rest of the CCN managers participate in the auction-based market. As can be seen from Figure 1, the proposed model involves two stages for gathering and selling the resources. The first stage captures the interactions between the PAs and the CCN managers, while the second represents the interactions between the CCN managers and the CPs. In the following, we will introduce and analyze these two stages in order.

3. INTERACTIONS BETWEEN CCN MANAGERS AND PAS: OPTIONS-BASEDSEQUENTIAL AUCTIONS

The main purpose of utilizing an auction is to sell goods when there is more than one interested buyers. In *sequential auction*, the seller holds consecutive auctions for selling goods. Since the seller can adjust the time interval between the consecutive auctions, sequential auctions are suitable for the following scenarios:

There are of two types:

(i) Availability of the goods varies over time, which means the goods may not be available in some of the time instances.

(ii) The buyers arrive at the market at different times, which requires the seller to wait for some period of time before the number of buyers exceeds a threshold to guarantee a certain profit it. Considering these facts, sequential auction is arguably the most suitable type of auctions for leasing the cloud servers to the PAs.

3.1First- and Second-Price OBSAs and PA's Role

Algorithm 1: Price matching process for the first-price options-based sequential auction

input : Current price that the PA has to pay when the current auction begins, entering and patience time of the winner PA ($P_{cur}, t_{ent}, t_{pat}$), clock time ($T$), winner of the current auction's bid (b_w)

output: Price that the PA has to pay ($P_{out}(T)$)

$t \longleftarrow t_{ent} + t_{pat}$
$P_{out}(T) \longleftarrow P_{cur}$
if ($t \geq T$ *and* $b_w < P_{cur}$) **then**
  $P_{cur} \longleftarrow b_w$
end
$P_{out}(T) \longleftarrow P_{cur}$

payment equal to his (the second highest) bid. Considering OBSAs in our context, the options-based property guarantees the least payment for winner PAs during their patience time, where the patience time is referred to as the time window in which the PA can wait before utilizing its obtained resources. Hence, this property eliminates the sensitivity of the PAs' payment to the time of winning an auction. In OBSAs, winner PAs are granted the opportunity to collect all of their demanded resources from the CCN resource pool before getting charged. The options-based property manifests itself through the price matching process, which is the main difference between OBSAs and classic sequential auctions. In [23], OBSAs with second-price backbone are proposed without mathematical analysis. In this work, we introduce them to the cloud-related literature, adapt them to the cloud resource allocation scenario, and provide mathematical analysis identifying multiple performance metrics of interest. We also present the OBSAs with the first-price backbone, which builds the foundation of analysis for the OBSAs with the second-

price backbone.

Algorithm 2: Price matching process for the second-price options-based sequential auction

input : Current price that the PA has to pay when the current auction begins, entering and patience time of the winner PA (P_{cur},t_{ent},t_{pat}), stored bumped PA's ID in this PA's memory (ID_{mem}), clock time (T), bid and ID of the current auction's winner (b_w,ID_w), bid and ID of the current auction's bumped PA (b_{bumped},ID_{bumped})
output: Price that the PA has to pay ($P_{out}(T)$)

```
t ⟵ t_ent + t_pat
P_out(T) ⟵ P_cur
if (t ≥ T) then
    if (ID_mem ≠ Null) then
        if ID_w = ID_mem then
            if b_bumped < P_cur then
                ID_mem ⟵ ID_bumped
                P_cur ⟵ b_bumped
            end
        else
            if b_w < P_cur then
                ID_mem ⟵ Null
                P_cur ⟵ b_w
            end
        end
    else
        if b_w < P_cur then
            P_cur ⟵ b_w
        end
    end
end
P_out(T) ⟵ P_cur
```

1) **Calculating the bid for each of the auctions of interest:** The PA submits a bid equal to the customer's *maximum marginal value* for suitable auctions [17]. By pursuing this bidding strategy, the PA wins those types of servers which are potentially profitable for the corresponding customer.

2) **Obtaining the best price (price matching):** There are two modes of operation for any PA in an auction: participant mode and observer mode. When a PA enters a CCN, he becomes a participant and participates in the appropriate active auctions. From the moment that the PA wins an auction, he switches to the observer mode for that auction. In this mode, the observer PA reduces the price of a won server to a lower price using the following procedure:

A)OBSAs with the first-price backbone: The observer PA decreases his current payment to the winner's bid if the winner wins the auction with a lower price as compared to the agent PA's current payment. Otherwise, the observer PA does not make any changes to his current payment. The price matching process for the first-price OBSA is summarized in Algorithm 1.

B) OBSAs with the second-price backbone: In this case, each PA has an identity that gets updated whenever he enters the market on behalf of a new customer. The winner PA stores the identity of the PA who has proposed the second highest bid, i.e., the so-called bumped PA. Possible situations for the subsequent auction and the corresponding actions of the observer PA's are as follows:

B-1)The bumped PA wins the next auction: The observer PA decreases his current payment to the second highest bid of the next auction and updates his memory by saving the identity of the PA who has proposed this bid.

B-2) The bumped PA stays at the market but loses the next auction: This implies that the winner's bid is higher than the current payment of the observer PA. This is due to the fact that the true valuation of the bumped PA is time invariant and that truthfulness is a dominant strategy in the second-price OBSAs [23]. In this case, the observer PA will neither change its memory nor its current payment.

B-3) The bumped PA leaves the market: The observer PA clears his memory and, from that point, he decreases his current payment to each of the successive winner's bid if it is lower than its current payment (same as the first-price backbone).The price matching process for the second-price OBSAs is summarized in Algorithm 2.

3) Utilizing the won servers: At the end of the PA's patience time, among his obtained (won) servers, he chooses those maximizing the corresponding that it contains customer's utility. Mathematically,

the PA chooses a set of servers obtained as S*= arg max [$v(s)$ $p(s)$], where S denotes the won servers by the PA during his patience time, and $p(s)$,$v(s)$ denote the payment and the customer's valuation corresponding to the servers belonging to set s, respectively. All the other acquired servers by this PA will be returned to the CNN without any charge.

As can be observed, in OBSA, the winner determination is very simple in both first- and second-price backbones since the winner for each auction is always the PA with the highest bid. Also, payment calculation is fairly simple through observing the current winner bid. In contrast, combinatorial auction, proposed in contemporary cloud-related literature (e.g., [18], [7]), requires solving the complex winner determination process and calculating complex payment functions for each PA.

ANALYSIS OF OBSAS

4.1 Backgrounds

The existence of the price matching process makes the analysis of OBSAs completely different from that of classic sequential auctions. In this section, mathematical analysis of OBSAs is presented for both first-price and second price backbones. Without loss of generality, the analysis is performed for an OBSA for one type of servers. It is assumed that the i^{th}PA's bid (bi) takes a discrete value such that bi $[v_{min}, v_{max}]$, £ i, where v_{max} v_{min}= $K\delta$, K £ Z+, and δ is the quantization size of the bids. In this case, *vmin* is the CCN manager's lowest affordable price (corresponding to zero profit) and v_{max} could be an upper bound obtained through market survey. Also, PAs' bids are assumed to be independent and identically distributed (i.i.d.) in each auction. It is further assumed that the average number of PAs in the market follows a Poisson distribution with mean

Φ, where, at each round of auction, Ω portion of them participate in the auction. Hence, the number of PAs for each auction follows a Poisson distribution with mean $\lambda = \Omega\Phi$. Our following analysis hold for an arbitrary distribution of the PA"s bids; however, specific bid distributions are needed to obtain more concrete results for further insights.

4.2 First-price OBSAs

4.2.1 Modeling the price matching process

We model the price matching process of an observer PA by a discrete time homogeneous Markov chain (DTHMC), where the state space consists of the possible payments of the PA. Initially, a winner PA enters into one of the states depending on his current payment. Afterward, this PA transits among different states of the DTHMC during his observation time, where his current state always represents his current payment during the price matching process. At the end of his observation time window, he leaves the DTHMC and makes a payment according to his last state. To carry out the analysis, we define the residual patience time Δ as the time duration in which the observer PA stays in theproposed DTHMC (i.e., the length of the observation mode). Figure 3 depicts the transition diagram of the proposed DTHMC, where L = (v_{max} v_{min})/δ, pij denotes the transition probability from state i to state j, and θj denotes the *stopping (leaving) probability* for state

j. The leaving probabilities indicate quitting the price matching process upon having zero residual patience time and making a payment with respect to the last state. Also, $\pi 0j$ denotes the *entering probability* for state *j*. This parameter is defined based on the probability of entering the DTHMC states for a winner PA which depends on the initial bids. Considering a lower bound on the bids reported by the CCN manager, all the states of the DTHMC are transient except for the last state. Also, due to the price matching process, only transitions to lower prices (states with smaller indices) are possible.

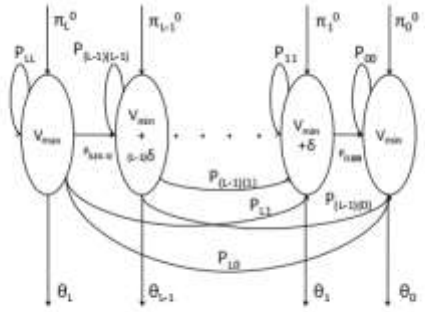

Figure 3: Transition diagram for the price matching process in the first-price OBSA.

4.3 Second-price OBSAs

4.3.1 Modeling the price matching process

Considering the price matching process of second price OBSAs presented in subsection 3.1, it can be seen that as long as the bumped PA stays in the market, the observer PA will only change his current payment when the bumped PA wins one of the subsequent auctions. This is because if the bumped PA stays in the market and does not win the subsequent auctions, the winner PA's bid will be higher than the observer PA's current payment. Nevertheless, when the bumped PA leaves the market, the observer PA clears his memory and adapts his payment to the winner PA's bid at each of the subsequent auctions. Similar to the first-price OBSAs, Figure 4 depicts the transition diagram of our proposed DTHMC corresponding to the actions of an observer PA in the second-price OBSAs. The transition diagram is composed of two main components:

A primary Markov chain and multiple chains. The primary Markov chain describes the case in which the bumped PA stays in the market, while the sub-Markov chains capture the actions upon leaving the bumped PA. Each *sub n*<*m* in Figure 4 refers to the case of leaving from the *mth* state of the primary Markov chain and entering a sub-Markov chain as depicted in Figure 3 with an initial state

$$n\ (\pi 0n = 1), \text{ with } n \leq m.$$

This corresponds to the case in which the current payment of the observer. It is assumed that the bumped PA participates in the next auction with probability $p(B)$, which is model as a complementary cdf of a memoryless distribution (e.g., exponential distribution) for the ease of analysis. According to the same logic used for deriving.

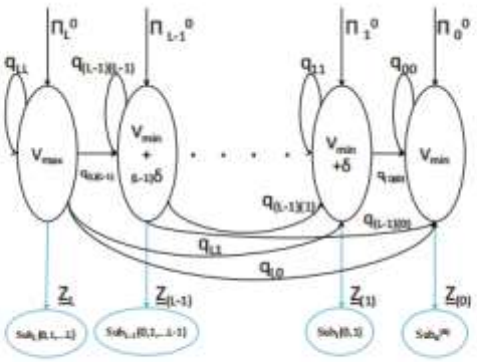

Figure 4: Transition diagram for the price matching process in the second-price OBSAs.

5. INTERACTIONS BETWEEN CCN MANAGERS AND CPS

In this section, we analyze the interactions between the CCN managers and the CPs. We consider a market in which CCN managers are the buyers and CPs are the sellers. The CCN managers compete to acquire resources from the CPs to full fill their demands and attract more customers. This is a dynamic market in the sense that both of the CCN managers and the CPs can leave and join the market over time.

For each type of servers, we divide the market into two sub-markets: flat-price market and auction-based market. In the flat-price market, resources are sold with a constant price. In the auction-based market, resources are sold through auctions. Consequently, each CCN manager has two options to obtain extra resources: (i) to buy the resources with a unit flat-price; (ii)to participate in the auctions. Correspondingly, the CPs have two options to sell their extra resources. In the rest of the discussion, without loss of generality, we focus on one type of servers.

It is assumed that auctions are held with a Poisson rate λA, and the number of CCN managers available in the market during each round of auction is Poisson distributed with mean λ *CCN*. In each auction, a portion of the CCN managers act as active bidders and participate in the auction, while the rest act as passive participants and do not submit their bids; the fraction of active CCN managers is denoted by μ. Moreover, we assume that the number of CPs participating in each round of the auction follows Poisson distribution with mean λCP . The time window in which a CCN manager can stay in the auction market (participation time) is limited to *Tp*. When the participation time of a CCN manager ends, he has to obtain the resources through the flat-price market with price *ß*.

In each auction, interested CCN managers participate with their bids reflecting their valuations for the servers, and CPs participate with their offered prices determining their least expected prices to lend their resources. The winner of each auction is the CCN manager who offers the highest bid, and he receives the resources from the CP with the lowest offered price. This happens if the lowest offered price is less than the winner's bid. Otherwise, there is no winner in the auction. As can be seen, two competitions emerge in this market: (i) the competition between CCN managers who try to offer higher prices to win the resources; (ii) the competition between CPs who try to offer lower prices to sell their resources. For the rest of this study, we focus on the CCN managers' competition since the competition among the CPs can be similarly analyzed. Consider one of the CCN managers with *r* units of residual participation time. Similar to [30], aggressive bidding strategy is considered, where the CCN managers bid more values as they approach the deadline. In this case, the instantaneous bid of a CCN manager can be model as

follows:

$$b(r) = e^{-\gamma r} u - D(r),$$

where u stands for the instant utility gained from the resource, γ the CCN manager's *rate of time preference*, and $D(r)$ the *discounted expected utility*. More precisely, $D(r)$ determines the cost of neglecting the potential future discounts by waiting in the auction. Note that $b(r)$ should be a decreasing function with respect to r since the CCN manager bids higher as he approaches the deadline. When a CCN manager's participation time ends, he has to leave the auction and pay the constant price ß for the resource. .This fact implies the following boundary conditions:

$$\mathbf{b(0) = ß,}$$

$$\mathbf{D(0) = u\ ß.}$$

It is assumed that the CCN managers' bids and the CP's offered prices are kept private. Hence, the CCN managers have no knowledge about the bids of each other. Nonetheless, the following information is assumed known to the CCN managers: (i) the cdf $F(b)$ and thus the pdf $f(b)$ of the bids; (ii) the cdf $G(o)$ and thus the pdf $g(o)$ of the CPs' offered prices.

As mentioned before, there are two well-known mechanisms for determining the winner's payment in the auction, first-price and second-price mechanisms. In-depth analysis of the both mechanisms in the described markets is presented below.

5.1 First-price Analysis

In the first-price payment scheme, the winner CCN manager has to pay the value of his bid to the seller CP. In the following theorem, we obtain the bidding strategy of a CCN manager in a stable market setting where the distributions of the bids and offered prices are stationary.

5.2 Second-price Analysis

In the second-price mechanism, the winner CCN manager has to pay the second highest bid to the seller CP. To obtain the CCN managers' bidding strategy in the stable market setting, we first derive the HJB equation for this case in the following theorem. Then, we propose a method to solve the derived HJB equation in the following corollaries.

6. SIMULATION RESULTS

The performance of the proposed scheme is illustrated through four simulation scenarios.

6.1 Scenario 1: CCN Managers' Income in OBSAs

We consider the market described in Table 1, where the bids' range is adopted from the history of winners' bid of Amazon's memory optimized instance (x1.32xlarge) in the past three month [4]. The minimum (maximum) value of bids corresponds to the 50% of the minimum winner bid (maximum winner bid) in that dataset upon requesting the cloud instance for 24 hours. The presented analytic results for the CCN manager's income in Section 4 and the results obtained from 10000 Monte Carlo simulations are depicted in Figure 5. It can be seen that the analytic results and the simulation results agree well with each other. Also it can be observed that the CCN manager's income becomes less sensitive to the number of PAs after the market gets sufficiently crowded. Moreover, the effect of increasing the patience time on the CCN manager's income becomes more pronounced when there are fewer participant PAs.

Auction mechanism	Bids' Range	Distribution	Parameters
Fist-price	[48,312]	Uniform	-
Second-price		Sampled Laplace	$\mu = 70, \delta = 1, w = 50$

Table 1: Simulation setting for the dynamic market between the PAs and the CCN managers.

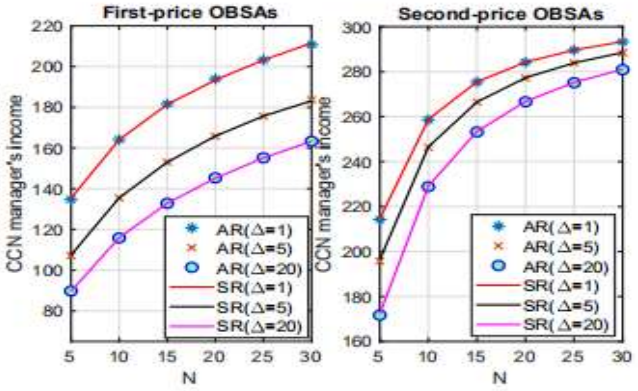

Figure 5: Analytic results (AR) and simulation results (SR) for the CCN manager's income in the first and second price OBSAs for various number of PAs (N) and different residual patience times (Δ).

6.2 Scenario 2: Bids of CCN Managers

To demonstrate the CCN managers' bidding behaviour in our proposed model, we simulate the result of Corollary 4. For this purpose, a market with the parameters described in Table 2 is considered. As before, considering the pricing history of Amazon's memory optimized instance in the past three month [4], the value of z is obtained assuming he maximum bid of a CCN for the cloud resources to be one-third of its maximum sold price. Due to the lack of a real dataset, the rest of the parameters are chosen based on common sense. Figure 6 depicts the value of the CCN managers' bids over time for various values of γ. As can be seen, as the value of γ increases, the CCN managers lose their interest in buying resources faster and bidding higher values.

Parameter	u	μ	λ_A	λ_{CCN}	λ_{CP}	z	a
Value	5	0.6	0.2	0.5	0.75	104	0.01

Table 2: Simulation setting for the dynamic market between the CCN managers and the CPs.

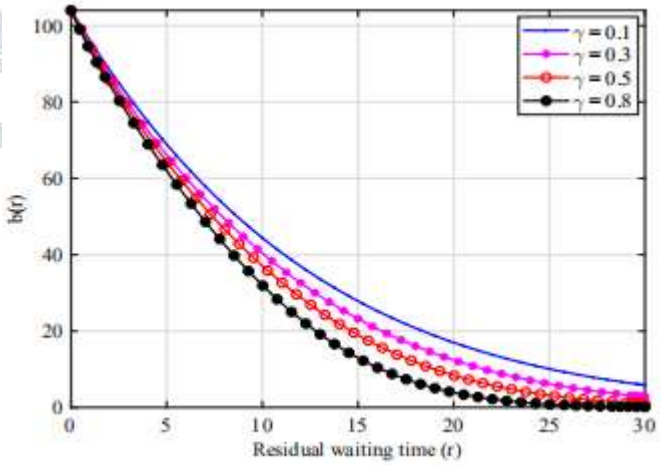

Figure 6: The CCN managers' bids with respect to the residual waiting time for different rates of time preference.

6.3 Scenario 3: Selling the Cloud Resources

In Figure 7, the improvement that can be obtained using the OBSAs instead of the sequential combinatorial auctions is shown in terms of the total number of winners and total sold cloud servers. Also, Figure 8 reveals the corresponding CCN manager's income. For this simulation, 10 types of servers are considered where the CCN manager holds 1000 auctions for each of these

types synchronously at every minute, where the total number of needed types of servers for each PA is uniformly distributed in the interval [1, 10]. The arrival rates (per minuet) for servers are i.i.d and uniformly distributed in the interval [1/15, 1]. This interval is derived based on the reported data in [19], which indicates that the majority of task executions are under 15 minutes. It is assumed that customers do not leave the market without obtaining their demanded resources (long patience time). The results in Figures 7, 8 indicate an increase in the customers' satisfaction, utilization of resources, and the CCN manager's income upon using the OBSAs.

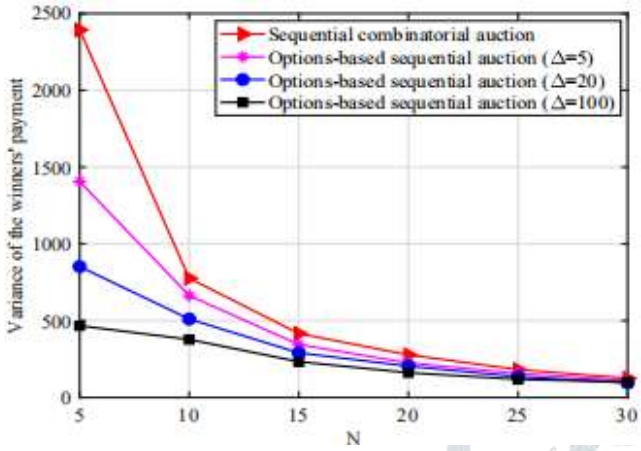

Figure 9: Comparison between the variance of the winners' payment in the second-price OBSA vs. sequential combinatorial auction for various number of PAs (N) and different residual patience times (Δ).

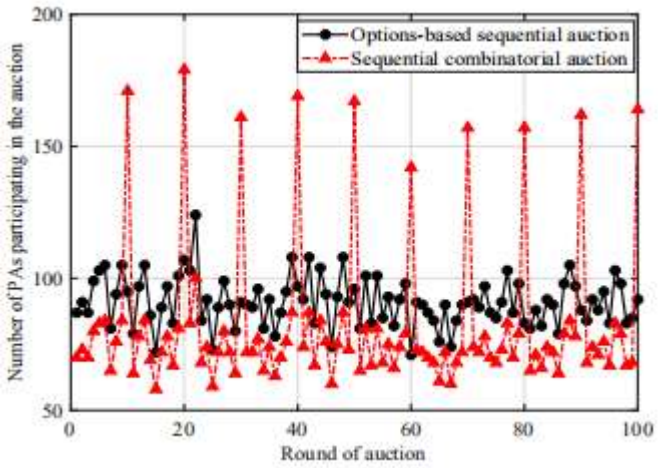

Figure 10: Comparison between the number of participant PAs in the second-price OBSA vs. sequential combination auction.

CONCLUSION AND FUTURE WORK

In this work, we proposed a two stage framework to describe resource allocation and gathering in modern cloud networks. The first stage describes the interactions between the PAs and the CCN managers. For this stage, OBSAs along with their theoretical analysis are proposed, which enjoy a simple winner determination process and provide the truthfulness property. The second stage models the interactions between the CCN managers and the CPs. For this stage, a theoretical framework is developed to model the bidding behaviour of the CCN managers. For future work, one worthwhile direction is to explore the optimization of the social welfare or other parameters of interest. Studying the resource allocation and the load balancing problems jointly is also interesting. In this case, a CCN manager should consider the geographical locations of the servers and CPs to find the optimal resource allocation.

REFERENCES

[1] Q. Wang, K. Ren, and X. Meng, "When cloud meets ebay: Towards effective pricing for cloud computing," in Proc. IEEE INFOCOM, March 2012, pp. 936–944.J.Carolan,S.Gaede,

[2] L. Gao, Y. Xu, and X. Wang, "MAP: Multiauctioneer progressive auction for dynamic spectrum access," IEEE Trans. Mobile Comput., vol. 10, no. 8, pp. 1144–1161, Aug 2011.

[3] Microsoft. (2017) Microst azure. [Online]. Available: https: //azure.microsoft.com/en-us/

[4] D. Bernhardt and D. Scoones, "A note on sequential auctions," American Econ. Rev., vol. 84, no. 3, pp. 653–657, 1994.

[5] Y. Zhang, D. Niyato, P. Wang, and E. Hossain, "Auction-based resource allocation in cognitive radio systems," IEEE Commun. Mag., vol. 50, no. 11, pp. 108–120, November 2012.

[6] U. Lampe, M. Siebenhaar, A. Papageorgiou, D. Schuller, and R. Steinmetz, "Maximizing cloud provider profifit from equilibrium price auctions," in Proc. IEEE 5th Int. Conf. Cloud Computer June 2012, pp. 83–90.

[7] W. Shi, L. Zhang, C. Wu, Z. Li, and F. C. M. Lau, "An online auction framework for dynamic resource provisioning in cloud computing," IEEE/ACM Trans. Netw., vol. 24, no. 4, pp. 2060 2073, Aug 2016.

[8] L. Jiang and S. Low, "Multi-period optimal energy procurement and demand response in smart grid with uncertain supply," in Proc. 50th IEEE Conf. Decision Control Eur. Control Conf. (CDC-ECC). IEEE, 2011, pp. 4348–4353.

[9] D. Coey, B. Larsen, and B. Platt, "A theory of discounts and Dead lines in retail search," Tech. Report, Working paper, Stanford University, June 2016.

[10] N. Nisan, "Bidding and allocation in combinatorial auctions," in Proc. 2nd ACM Conf. Electron. Commerce, New York, NY, USA, 2000, pp. 1–12.

[11] D. Bernstein, E. Ludvigson, K. Sankar, S. Diamond, and M. Morrow, "Blueprint for the intercloud - protocols and formats for cloud Computing interoperability," in Proc. 4th Int. Conf. Internet Web Appl. Serv., May 2009, pp. 328–336.

[12] X. Wang, Z. Li, P. Xu, Y. Xu, X. Gao, and H. H. Chen, "Spectrum sharing in cognitive radio networks-An auction- based approach," IEEE Trans. Syst., Man, Cybern., Part B (Cybernetics), vol. 40, no. 3, pp. 587–596, June 2010.

[13] C. Boutilier, M. Goldszmidt, and B. Sabata, "Sequential auctions

for the allocation of resources with complementarities," in Proc. 16th Int. Joint Conf. Artifificial Intell., vol. 99, 1999, pp. 527–523.

[14] H. Zhang, H. Jiang, B. Li, F. Liu, A. V. Vasilakos, and J. Liu, "A framework for truthful online auctions in cloud computing with heterogeneous user demands," IEEE Trans. Comput., vol. 65, no. 3, pp. 805–818, March 2016.

[15] S. Hosseinalipour and H. Dai, "Options-based sequential auctions for dynamic cloud resource allocation," in Proc. IEEE Int. Conf. Commun. (ICC), May 2017, pp. 1–6.

[16] W. Wang, B. Liang, and B. Li, "Revenue maximization with dy namic auctions in iaas cloud markets," in Proc. IEEE/ACM 21st Int.

[17] Symp. Quality Serv. (IWQoS), June 2013, pp. 1–6.

A. I. Juda and D. C. Parkes, "An options-based solution to the sequential auction problem," Artifificial Intell., vol. 173, no. 7-8, pp. 876–899, 2009.

[18] S. Zaman and D. Grosu, "A combinatorial auction-based mech anism for dynamic VM provisioning and allocation in clouds," IEEE Trans. Cloud Comput., vol. 1, no. 2, pp. 129–141, July 2013.

[19] C. Reiss, A. Tumanov, G. R. Ganger, R. H. Katz, and M. A. Kozuch, "Towards understanding heterogeneous clouds at scale: Google trace analysis," Tech. Report, Intel Science and Technology Center for Cloud Computing, 2012.

[20] L. Gao, X. Wang, Y. Xu, and Q. Zhang, "Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach," IEEE J. Sel. Areas Commun., vol. 29, no. 4, pp. 843–855, April 2011.

[21] Google. (2017) Google cloud platform. [Online]. Available: https://cloud.google.com/

[22] K. M. Sim, "Agent-based interactions and economic encounters in an intelligent intercloud," IEEE Trans. Cloud Comput., vol. 3, no. 3, pp. 358–371, July 2015.

[23] J. R. Norris, Markov Chains. Cambridge University Press, 1998.

[24] C.Boutilier and H. H. Hoos, "Bidding languages for combinatorial auctions," in Proc. 17th Int. Joint Conf. Artifificial Intell., 2001, pp. 1211–1217.

[25] N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani, Algorithmic game theory. Cambridge University Press, 2007, vol.

[26] L. Zhang, Z. Li, and C. Wu, "Dynamic resource provisioning in cloud computing: A randomized auction approach," in Proc. IEEE INFOCOM, April 2014, pp. 433–441.

[27] A. W. Min, X. Zhang, J. Choi, and K. G. Shin, "Exploiting spectrum heterogeneity in dynamic spectrum market," IEEE Trans. Mobile Comput., vol. 11, no. 12, pp. 2020–2032, Dec 2012.

[28] Amazon. (2017) Amazon elastic compute cloud (amazon EC2). [Online]. Available: https://aws.amazon.com/ec2/

[29] V. P. G, S. Rao, and A. S. Prasad, "A combinatorial auction mechanism for multiple resource procurement in cloud computing," inProc. 12th Int. Conf. Intell. Syst. Design Appl., Nov 2012, pp. 337–344.

[30] H. Zhang, B. Li, H. Jiang, F. Liu, A. V. Vasilakos, and J. Liu, "A framework for truthful online auctions in cloud computing with heterogeneous user demands," in Proc. IEEE INFOCOM, April 2013, pp. 1510–1518.

[31] D. Bernhardt and D. Scoones, "A note on sequential auctions," American Econ. Rev., vol. 84, no. 3, pp. 653–657, 1994.

[32] Amazon. (2017) Amazon ec2 spot instances. [Online]. Available: https://aws.amazon.com/ec2/spot/.

Seyyedali Hosseinalipour (S'17) received the B.S. degree in Electrical Engineering from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran in 2015. He is currently pursuing a Ph.D. degree in the Department of Electrical and Computer Engineering at North Carolina State University, Raleigh, NC, USA. His research interests include analysis of wireless networks, resource allocation and load balancing for cloud networks and resource allocation and task scheduling for vehicular ad-hoc networks.

Huaiyu Dai (F'17) received the B.E. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ in 2002.

He was with Bell Labs, Lucent Technologies, Holmdel, NJ, in summer 2000, and with AT&T Labs-Research, Middletown, NJ, in summer 2001. He is currently a Professor of Electrical and Computer Engineering with NC State University, Raleigh. His research interests are in the general areas of communication systems and networks, advanced signal processing for digital communications, and communication theory and information theory. His current research focuses on networked information processing and crosslayer design in wireless networks, cognitive radio networks, network security, and associated information-theoretic and computation-theoretic analysis.