AUCTION BASED RESOURCE ALLOCATION FOR CLOUD SERVICE MANAGEMENT SYSTEM

B.Rama[1], V.Usha Nanthini[2] , R.Lavanya[3]
[1]. Associate Professor, Department of IT , M.A.M.College of Engineering and Technology,Trichy
[2]. Student, Department of IT, M.A.M.College of Engineering and Technology,Trichy
[3]. Student, Department of IT, M.A.M.College of Engineering and Technology,Trichy

Abstract--The contemporary literature on cloud resource allocation is mostly focused on studying the interactions between customers and cloud managers. Nevertheless, the recent growth in the customers' demands and the emergence of private cloud providers (CPs) entice the cloud managers to rent extra resources from the CPs so as to handle their backlogged tasks and attract more customers. This also renders the interactions between the cloud managers and the CPs an important problem to study. In this paper, we investigate both interactions through a two-stage auction mechanism. For the interactions between customers and cloud managers, we adopt the options-based sequential auctions (OBSAs) to design the cloud resource allocation paradigm. As compared to existing works, our framework can handle customers with heterogeneous demands, provide truthfulness as the dominant strategy, enjoy a simple winner determination procedure, and preclude the delayed entrance issue. We also provide the performance analysis of the OBSAs, which is among the first in literature. Regarding the interactions between cloud managers and CPs, we propose two parallel markets for resource gathering. We capture the selfishness of the CPs by their offered prices. We conduct a comprehensive analysis of the two markets and identify the bidding strategies of the cloud managers.

Index Terms—Auction theory, cloud of clouds networks, sequential auctions, options-based sequential auctions, proxy agent, cloud resource allocation, Hamilton-Jacobi-Bellman equation, dynamic market

1. INTRODUCTION

Modern society relies crucially on efficient processing of the massive amount of data collected from a variety of sources such as wireless sensors and statistical polls, for which cloud computing is a natural platform. Various cloud-based services have been offered, including Microsoft Azure [2], Google Cloud [3], and Amazon EC2 [4], while many other companies are planning to join this profitable market. The recent growth in the customers' demands has motivated the idea of resource sharing in cloud networks [5], 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 is the auction mechanism due to its simplicity and versatility, which is 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 [6].

2. LITERATURE REVIEW

The Objective of this paper is to investigate the two correspondences through a two-mastermind auction framework. For the correspondences among customers and Product Owners, we get the OBSA technique to structure the resource task perspective.

1. TITLE: Combinatorial Reverse Auction based Scheduling in Multi-Rate Wireless Systems

Opportunistic scheduling algorithms are effective in exploiting channel variations and maximizing system throughput in multirate wireless networks. However, most scheduling algorithms ignore the per-user quality-of-

service (QoS) requirements and try to allocate resources (for example, the time slots) among multiple users. This leads to a phenomenon commonly referred to as the exposure problem, wherein the algorithms fail to satisfy the minimum slot requirements of the users due to substitutability and complementarity requirements of user slots. To eliminate this exposure problem, we propose a novel scheduling algorithm based on two-phase combinatorial reverse auction, with the primary objective of maximizing the number of satisfied users in the system.

This paper also consider maximizing the system throughput as a secondary objective. In the proposed scheme, multiple users bid for the required number of time slots and the allocations are done to satisfy the two objectives in a sequential manner. The author provide an approximate solution to the proposed scheduling problem, which is NP-complete. The proposed algorithm has an approximation ratio of (1 + log m) with respect to the optimal solution, where m is the number of slots in a schedule cycle. Simulation results are provided to compare the proposed scheduling algorithm with other competitive scheme.

2. TITLE: Two Phase Scheduling Algorithm for Maximizing the Number of Satisfied Users in Multi-Rate Wireless Systems

Opportunistic scheduling algorithms are effective in exploiting channel variations and maximizing system throughput in multi-rate wireless networks. However, most scheduling algorithms ignore the per-user quality of service (QoS) requirements and try to allocate resources (i.e., the time slots) among multiple users. This leads to a phenomenon commonly referred to as the exposure problem wherein the algorithms fail to satisfy the minimum slot requirements of the users due to substitutability and complementarity requirement of user slots. To eliminate this exposure problem, we propose a novel scheduling algorithm based on two phase combinatorial reverse auction with the primary objective to maximize the number of satisfied users in the system. It also consider maximizing the system throughput as a secondary objective. In the proposed scheme, multiple users bid to acquire the required number of time slots, and the allocations are done to satisfy the two objectives in a sequential manner. We provide an approximate solution to the proposed scheduling problem which is a NP-complete problem. We prove that our proposed algorithm is (1 + log m) times the optimal solution, where m is the number of slots in a schedule cycle. This author also present an extension to this algorithm which can support more satisfied users at the cost of additional complexity. Numerical results are provided to compare the proposed scheduling algorithms with other competitive schemes.

3. TITLE: Spectrum Trading In Cognitive Radio Network a Contract Theoretic Model Approach

Dynamic spectrum allocation via auction is an effective solution to spectrum shortage. Combinatorial spectrum auction enables buyers to express diversified preferences towards different combinations of channels. Despite the effort to ensure truthfulness and maximize social welfare, spectrum auction also faces potential security risks. The leakage of sensitive information such as true valuation and location of bidders may incur severe economic damage. However, there is a lack of works that can provide sufficient protection against such security risks in combinatorial spectrum auction. In this paper, we propose ARMOR, to enable combinatorial auction for heterogeneous spectrum with privacy, which can preserve bidders' privacy while guaranteeing the economic-robustness of the combinatorial auction. We leverage the cryptographic methods, including homomorphic encryption, order-preserving encryption, and garbled circuits, to shield the bid and location information of buyers from the auctioneer. We design a novel location protection algorithm, which allows the auctioneer to exploit spectrum reuse opportunities without knowing the exact locations of buyers. Furthermore, we propose a verifiable payment scheme based on digital signature to prevent the auctioneer from forging the payment. The extensive experiments confirm that ARMOR maintains the good performance of the combinatorial spectrum auction, in terms of buyer satisfactory ratio and social welfare, and achieves privacy preservation with acceptable computation and communication costs.

4. TITLE: Auction-Based Resource Allocation in Cognitive Radio Systems

Auction theory, as a subfield of economics, provides useful tools to model, analyze, and optimize radio resource management in cognitive radio environments. By using an auction, radio resources such as sub channel, time slot, and transmit power can be allocated

among licensed and unlicensed users in the system, following market laws. Due to the flexibility of mechanism design, there are various auction mechanisms that have been applied to cognitive radio systems with different characteristics. In this article, we first provide an overview of the basics of general auctions. Then the motivations and specific design issues in applying auctions to wireless network architectures and protocols are discussed. Then we review the state of the art in the use of auction theory and mechanism design in cognitive radio networks. This will enable the readers to have a general view of auction fundamentals, as well as the recent development and applications of auction theory in the emerging cognitive wireless networks.

5. TITLE: An Auction Mechanism for Product Verification using Cloud

The recent growth in the cloud computing resource had witnessed demands for the cloud server in the public and private sector companies. To meet these demands, the cloud resource is rented on the basis of auction mechanism due to its simplicity and versatility. In this paper, we are presenting the allocation of a commodity to the stakeholders using the bidding process in the cloud server. Under this auction, entity such as customer (buyer), owner (seller) participate, but to enhance the trustfulness of the bidding a novelty (Government) is added handling the entire auction process. The auction is performed using Optional Based Sequential auction (OBSA) algorithm in two stages (i-e) verification of commodity and price matching function. At the end, the verified commodity with highest price quoted is sold to the customer.

3. MARKET STABILITY

To compares the variance of the winners' payment for the second-price OBSA described in Scenario 1 and that of the sequential combinatorial auction for various number of PAs (N) and different residual patience times (Δ). Less variance for the payments in the former is seen in Figure 9 implying a more predictable market for the CCN manager. Also, similar results are observed for the first-price OBSAs, which are omitted for the interest of space. Figure 10 depicts the number of participant PAs in the second-price OBSA and that in the sequential combinatorial auction. In this scenario, at each time instant one auction occurs, where the number of available PAs at each time instant is assumed to be a Poisson random variable with mean 93. This value is adopted from assuming that 10% of all the PAs participate in each round of auction. Each PA can delay his participation. It is assumed that 20% of these PAs have side information about the future market situation. It is assumed that the market receives the lowest bids from participant PAs in every 10 time instances. As can be seen, by utilizing sequential combinatorial auctions, those informed PAs delay their entrance into the market, leading to an unfavorable burst of arrival. In contrast, this issue is effectively suppressed by the proposed OBSAs.

4. METHODALOGY

Auction theory provides a solid mathematical foundation for resource allocation among a set of resource-seeking customers and a set of resource providers. There exists a body of literature studying auction-based resource allocation in other contexts such as spectrum sharing in cognitive radio networks [7], [8], [9], [10], and [11].

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 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 auctions and seek truthful sequential auction solutions.

The most related works to ours are 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. 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 it's 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 mainly designed for one-round selling. These issues of combinatorial auctions have promoted further research on solving winner determination using simpler approximation methods or extending them to sequential combinatorial auctions. In, the authors proposed a truthful mechanism for sequential combinatorial auctions. In this framework, besides complicated winner determination and payment identification processes, 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 the interactions between customers and cloud providers are modeled as an online combinatorial auction. The model of this 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.

However, similar to 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 aforementioned 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 a general framework for inter cloud networks is presented where the interactions between users and cloud providers are modeled by many-to-many auctions. Afterward, the interactions between cloud providers are modeled by a coalition game in which the cloud providers borrow resources from each other to fulfill their customers' 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 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 this scenario, all the users with side-information postpone their entrance into the market. In a large-scale market, this circumstance leads to burst arrivals, and thus an unstable market. In summary, existing literature on modeling the interactions between the cloud managers and customers has at least one of the following four limitations: incapability of handling customers' heterogeneous demands that require a bundle of different types of servers, missing the truthfulness property, requiring prohibitive computation for winner and payment determination, and susceptibility to the delayed entrance issue. These issues will bead dressed 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 through their different offered prices.

Novelty and Contributions

To address the limitations of the existing works me, and novel two-stage auction framework is

prop[11] 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. [12] W. Wang, B. Liang, and B. Li, "Revenue maximization with dynamic auctions in iaas cloud markets," in Proc.IEEE/ACM21stInt. Symp. Quality Serv. (IWQoS), June 2013, pp. 1–6.used in this work to capture the interactions among (a) 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 heterogeneous 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 enlarge 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) 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 provide the corresponding performance analysis is 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.3 We also provide a comprehensive analysis using Hamilton-Jacobi Bellman equations and derive the bidding strategies of the CCN managers with respect to their inherent characteristics in a stable market setting.

Structure of the paper:

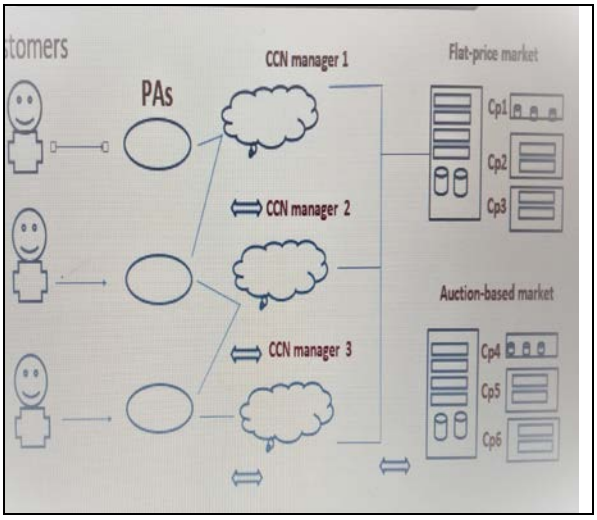
The system model is introduced in the interactions between the customers and the CCN managers are modeled

1. In this study, we consider the infrastructure as a service (IaaS) for of cloud computing.
2. The truthfulness property is guaranteed in the second-price options-based sequential auctions.
3. These markets may be viewed as the counter arts of the day-ahead and real-time markets in smart grids.
4. Devoted to the analysis of OBSAs. The interactions between the CPs and the CCN managers are modeled and analyzed. Simulation results are presented it concludes the paper and provides some possible future directions.

3. SYSTEM MODEL

A CCN consists of multiple cloud servers with different processing capabilities; some of them are more desirable for GPU processing, while the

ARCHITECTURE DIAGRAM:



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 back logged 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 proxy agents (PA) are incorporated into our model as trusted mediators between the customers and the corresponding CCN. 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 therein for more details. Also, there is a recent trend of research on designing privacy preserving auctions for spectrum allocation in wireless networks, these works focus on adapting effective cryptographic methods to combinatorial auctions and double auctions to protect location information and bidding values of the bidders. In this work, we assume that the bidding values and the private information of the customers are directly shared with the corresponding PAs. Protecting these information calls for adapting effective cryptographic methods to our proposed auction scheme, which is left as a future work. 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 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 the 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 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 these represents the interactions between the CCN managers and the CPs. In the following, we will discuss and analyze these two stages in order.

3.1 INTERACTIONS BETWEEN CCN MANAGERS AND PAS: OPTIONS-BASED SEQUENTIAL 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: (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. Considering these facts, sequential auction is arguably the most suitable type of auctions for leasing the cloud servers to the PAs. Classic first-price and second-price sequential auctions have been studied in the literature. Nevertheless, one of the main drawbacks of these auctions is the lack of a dominant strategy that can accommodate heterogeneous demands of buyers when customers face multiple sequential auctions. For example, consider the following two situations: 1) A buyer with a limited budget and heterogeneous demands requires goods from either of the two sequential auctions but not from both. 2) A buyer with a limited budget requires goods from both sequential auctions simultaneously. In both cases, a buyer has no dominant strategy for splitting the budget between multiple sequential auctions. For the interactions between the PAs and the CCN managers, considering the Pas as buyers, the and each server as a good, holding a separate sequential auction for each type of servers will lead to the situations.

In the above-mentioned examples, which makes classic sequential auctions inapplicable. Another main concern of applying the classic sequential auctions and the combinatorial auctions to our context is the delayed entrance issue, where PAs may intentionally delay their entrance to the market for a better price. This phenomenon leads to an undesired burst of arrivals into the market and causes market instability. The aforementioned concerns

suggests a need for an auction mechanism that can:

(i) Capture the dynamic is of the market with its sequential style of holding;

(ii) Enjoy truthfulness as the dominant strategy to ensure that PAs cannot make more profit by manipulating their true valuations;

(iii) Resolve the delayed entrance issue. To this end, we propose utilizing the options-based sequential auction (OBSA) in our study. Besides enjoying all the above motioned characteristics, OBSA leads to more trust between bidders and an auction, and thus result sine higher long-term profit for the auctioneer. Moreover, OBSA shave simple implementation and admit fast winner recognition. Basically, OBSAs are classic sequential auctions reinforced with the options-based property. Consider a classic sequential auction with first- or second-price backbone in our context. In the first-(second-) price scenario, the PA with the highest bid is the winner who is then required to make a 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 the time window in which the PA can wait before utilizing its obtained resources. Hence, this property eliminates the sensitivity of the PA's 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 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

CONCLUSION AND FUTURE WORK

In this work, we have proposed a comprehensive 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 behavior of the CCN managers. For future work, one 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] 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.

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

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

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

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

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

[7] 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.

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

[9] L. Gao, Y. Xu, and X. Wang, "MAP: Multi auctioneer progressive auction for dynamic

spectrum access, IEEE Trans .Mobile Comput" vol. 10, no. 8, pp. 1144–1161, Aug 2011.

[10] 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.

[11] 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.

[12] W. Wang, B. Liang, and B. Li, "Revenue maximization with dynamic auctions in iaas cloud markets," in Proc.IEEE/ACM21stInt. Symp. Quality Serv. (IWQoS), June 2013, pp. 1–6