# A Market for Scheduling, with Applications to Cloud Computing

Sadra Yazdanbod
Advisor: Vijay Vazirani

November 10, 2015

**Introduction:** The Internet has transformed not only the economy but also the central object of study of economists, the market, by creating innovative and immensely important marketplaces. The most important such market so far was the multi-billion dollar adwords market, and researchers from algorithms and AGT have contributed handsomely to its efficient operation [8, 3, 5, 7, 9, 6, 1]. A quickly emerging market is the cloud resources market. Since most projections predict that this market will dwarf even the adwords market, it is quintessential to understand its idiosyncrasies and design algorithms and mechanisms for its efficient operation.

The power of the pricing mechanism is well explored and understood in economics: It allocates resources efficiently since prices send strong signals about what is wanted and what is not, and it prevents artificial scarcity of goods while at the same time ensuring that goods that are truly scarce are conserved. Hence it is beneficial to both consumers and producers. For this reason, we propose an equilibrium-based mechanisms for cloud resources market. In our proposed model, the resources, such as computing power and memory, come in time slots and the agents have specific requirements and would like to buy resources in the earliest time slots subject to their budget. The problem is to compute an equilibrium for this market, i.e., prices at which demand equals supply for each resource. I will describe the details in the next section.

Two important aspects required in this model are: (1) agents desire *duration guarantees*: they must be allocated each resource for a specified amount of time, and (2) agents would like to complete their jobs as soon as possible. In incorporating these aspects, we arrive at a model that deviates substantially from market models studied so far in economics and theoretical computer science. The latter models define preferences of agents via a utility function: an agent prefers that bundle of goods which maximizes a given (usually concave) utility function, subject to a budget constraint; strict constraints on goods are never allowed. In contrast, our model allows agents to state their exact requirement for each good, and satisfies these covering requirements fully. In addition, our model incorporates a temporal aspect, i.e., there is a notion of time in the model and prices of resources are a function of time. These features make our market model suitable for resource allocation and scheduling, which are central to cloud computing.

**Our Model:** Let $A$ be a set of $n$ agents, indexed by $i$, and $G$ be a set of $m$ different types of goods, indexed by $j$. Time is divided into *slots*, each of one time unit. Assume that there are $s$ of slots and they are indexed by $t$. Let $x_{ijt}$ be the allocation of good $j$ to agent $i$ in slot $t$. Each agent $i$ has a set of *jobs*. Each job $k$ is defined by a continuous, non-decreasing, and concave, *rate function* , $U_i^k$. When given a bundle of goods $\boldsymbol{x}_{it}^k \in \mathbb{R}_+^G$ in slot $t$, a $U_i^k(\boldsymbol{x}_{it}^k)$ fraction of job $k$ that can be completed in that slot. Therefore the requirement for this job is $\sum_t U_i^k(\boldsymbol{x}_{it}^k) \geq 1$. The agent wants to complete all her jobs at a cost no larger than her budget $m_i$, *i.e.,* $\sum_k \sum_t \boldsymbol{x}_{it}^k \cdot \boldsymbol{p}_t \leq m_i$, and minimize the weighted flow time, $\sum_k w_i^k \sum_t t U_i^k(\boldsymbol{x}_{it}^k)$. Market clears when there is no over demand of any good, and demand $<$ supply implies the price is 0.

**Our Results:** First question that comes to mind here is that whether there even exists such equilibrium prices. Interestingly enough we manage to show that there always exists an equilibrium assuming a few mild assumptions. However our proof just gives us the existence and the problem is whether we can

efficiently find such prices.

Given the strong connection between equilibria in Fisher markets and convex programs, one might ask whether there exists a convex program that captures the equilibrium allocation/prices in this market. We show via examples that this is unlikely to be the case. In fact, in addition to non-convexity, the set of equilibrium prices could have "holes" in between, i.e., this set could have genus $> 0$. This immediately rules out any convex programming based solution. All other known techniques developed to compute equilibria in markets in the last decade and half don't seem to apply here. Therefore, there is a need for developing new techniques.

In our recent work [4], we study a special case of our general modal where each agent $i$ has a requirement of $r_{ij}$ ($\geq 0$) units of each good $j$. Agents seek to buy resources at the earliest time possible subject to their budget to satisfy their requirement, i.e., $\sum_t f_{ijt} \geq r_{ij}$, $\forall j$. It is worth noting that the discussion we had in the introduction holds for this case as well. We manage to provide a polynomial time algorithm to compute an equilibrium in this market using a new technique. We observe that the right thing to do here is to search in the space of price slopes, or equivalently exchange rates between cost and delay for each agent. It turns out that the optimal way for an agent to spread his money among different goods is to combine delay and price linearly using a coefficient $\lambda_i$ for delay (and 1 for price), and buy the cheapest $r_{ij}$ slots of good $j$ according to this. Agents who have a higher $\lambda_i$ get scheduled earlier. In each iteration of the algorithm, we identify the set of buyers who are going to have the smallest $\lambda_i$s, and get the latest of the slots. We show that we can set the prices of these slots so that it doesn't conflict with our future choices, and recurse with the remaining buyers.

**Future Direction:** There are many interesting special cases of our general model, and we expect that this model will provide a rich set of questions regarding computability of equilibria, and otherwise. As a very next step, we want to consider the case where each job requires a certain combination of resources for a certain amount of time to complete. For example, suppose there are two types of resources available like computing power and memory. Then each job needs a certain combination of computing power and memory for a certain amount of time for completion. Since this is a special case of our general model we have the existence of a price equilibrium. But the question we want to answer is that whether it can be found efficiently. We hope to attack this problem by extending our techniques in [4].

## Predicting Agents' Actions in a Repeated Market Model

**Introduction:** Companies like Google run billions of auctions every day to sell ads. Any change to the rules of this enormous repeated market can have a tremendous effect on the revenue of the company and the welfare of the agents. Therefore, any change requires careful evaluation of its potential impact. Currently, such impact is evaluated by running controlled experiments, which can only be done on a very small fraction of the daily traffic. As a result, this misses the important fact that the advertisers might act completely differently if the changes are applied to the whole traffic. Our goal is to build a theoretical framework for predicting advertiser response based on different rationalities of the advertisers.

**Our Model:** We model this problem using the multi-armed bandit setting [2], where each arm corresponds to a strategy. The cost of these arms change over time and are publicly observable. The value of playing an arm is drawn stochastically from a static distribution and is observed by the advertiser but not by us. We, however, observe the actions of the advertiser. Assuming the advertiser is playing a strategy with a regret of at most $f(T)$ within the first $T$ rounds, we want to predict the advertisers actions in such a way that the regret of our predictions is bounded by something as close as possible to $f(T)$.

We conjecture that there exists an efficient algorithm with a prediction regret bound of at most $O(f(T))$. So far we have found an efficent algorithm with a prediction bound of $O(f(T)\ln(T))$.

## References

[1] G. Aggarwal, A. Goel, and R. Motwani. Truthful auctions for pricing search keywords. In *Proceedings of the 7th ACM conference on Electronic commerce*, pages 1–7. ACM, 2006.

[2] S. Alaei, M. Mahdian, A. B. Varadaraja, and S. Yazdanbod. Predicting advertiser response. *In preparation.*

[3] N. Buchbinder, K. Jain, and J. S. Naor. Online primal-dual algorithms for maximizing ad-auctions revenue. In *Algorithms–ESA 2007*, pages 253–264. Springer, 2007.

[4] N. R. Devanur, J. Garg, R. Mehta, V. V. Vazirani, and S. Yazdanbod. A market for scheduling, with applications to cloud computing. *Submitted to the 48th ACM Symposium on Theory of Computing (STOC 2016).*

[5] N. R. Devanur and T. P. Hayes. The adwords problem: online keyword matching with budgeted bidders under random permutations. In *Proceedings of the 10th ACM conference on Electronic commerce*, pages 71–78. ACM, 2009.

[6] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. Technical report, National Bureau of Economic Research, 2005.

[7] J. Feldman, A. Mehta, V. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 117–126. IEEE, 2009.

[8] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani. Adwords and generalized online matching. *Journal of the ACM (JACM)*, 54(5):22, 2007.

[9] H. R. Varian. Position auctions. *international Journal of industrial Organization*, 25(6):1163–1178, 2007.