Bidding Dynamics of Rational Advertisers in Sponsored Search Auctions on the Web

S. Siva Sankar Reddy and Y. Narahari
Electronic Commerce Laboratory
Department of Computer Science and Automation
Indian Institute of Science, Bangalore - 560 012
sankar,hari@csa.iisc.ernet.in

Abstract— In this paper, we address a key problem faced by advertisers in sponsored search auctions on the web: how much to bid, given the bids of the other advertisers, so as to maximize individual payoffs? Assuming the generalized second price auction as the auction mechanism, we formulate this problem in the framework of an infinite horizon alternative-move game of advertiser bidding behavior. For a sponsored search auction involving two advertisers, we characterize all the pure strategy and mixed strategy Nash equilibria. We also prove that the bid prices will lead to a Nash equilibrium, if the advertisers follow a myopic best response bidding strategy. Following this, we investigate the bidding behavior of the advertisers if they use Q-learning. We discover empirically an interesting trend that the Q-values converge even if both the advertisers learn simultaneously.

Keywords

Mechanism Design, Internet Advertising, Sponsored Search Auctions, GSP (Generalized Second Price) Auction, Myopic Best Response Bidding, Q-Learning.

I. INTRODUCTION

A. Sponsored Search Auctions

Figure 1 depicts the result of a search performed on Google using the keyword '*auctions*'. There are two different stacks - the left stack contains the links that are most relevant to the query term and the right stack contains the sponsored links. Sometimes, a few sponsored links are placed on top of the search search results as shown in the Figure I-A. Typically, a number of merchants (advertisers) are interested in advertising alongside the search results of a keyword. However, the number of slots available to display the sponsored links is limited. Therefore, against every search performed by the user, the search engine faces the problem of matching the advertisers to the slots. In addition, the search engine also needs to decide on a price to be charged to each advertiser. Note that each advertiser has different desirability for different slots on the search result page. The visibility of an Ad shown at the top of the page is much better than an Ad shown at the bottom and, therefore, it is more likely to be clicked by the user. Therefore, an advertiser naturally prefers a slot with higher visibility. Hence, search engines need a system for allocating the slots to advertisers and deciding on a price to be charged to each advertiser.

Due to increasing demands for advertising space, most search engines are currently using auction mechanisms for
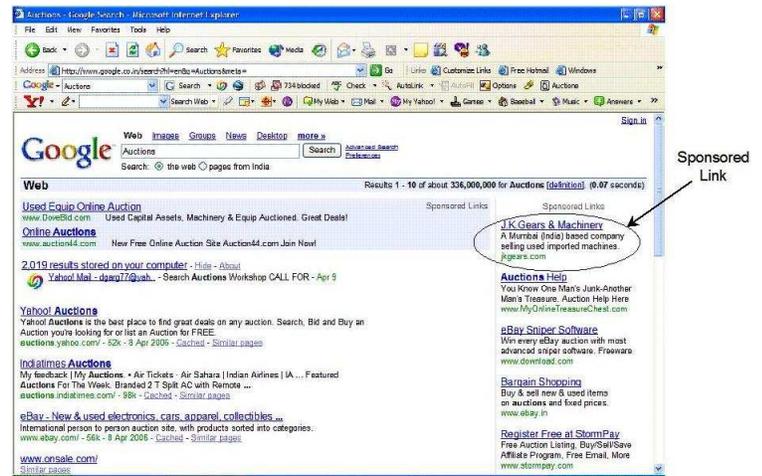

Fig. 1. Result of a search performed on Google

this purpose. In a typical sponsored search auction, advertisers are invited to submit bids on keywords, i.e. the maximum amount they are willing to pay for an Internet user clicking on the advertisement. This is typically referred by the term *cost-per-click*. Based on the bids submitted by the advertisers for a particular keyword, the search engine (which we will sometimes refer to as the auctioneer or the seller) picks a subset of advertisements along with the order in which to display. The actual price charged also depends on the bids submitted by the advertisers. There are many terms currently used in practice to refer to these auctions models, e.g. *Internet search auctions, sponsored search auctions, paid search auctions, paid placement auctions, AdWord auctions, Ad Auctions, slot auctions, etc.*

Currently, the auction mechanisms most widely used by the search engines are based on GSP (Generalized Second Price Auction) [1], [2], [3]. In GSP auctions, each advertiser would bid his/her maximum willingness to pay for a set of keywords, together with limits specifying his/her maximum daily budget. Whenever a keyword arrives, first the set of advertisers who have bid for the keyword and having non-zero remaining daily budget, are determined. Then, the Ad of

advertiser with the highest bid would be displayed in the first position; the Ad of advertiser with the next highest bid would be displayed in the second position and so on. Subsequently whenever a user clicks an Ad at position k, the search engine charges the corresponding advertiser an amount equal to the next highest bid i.e. the bid of advertiser allocated to the $(k+1)^{th}$ position.

B. Background for the Problem

The GSP mechanism does not have an equilibrium that is *incentive compatible* (that is, bidding their true valuations does not constitute an equilibrium). This leads to strategic bidding by the advertisers involving escalating and collapsing phases [1]. In practice, it is not clear whether there is a straightforward strategy for advertisers to follow. In fact, it is said that the overwhelming majority of marketers have misguided bidding strategies, if they have a strategy at all. This is surprising, given the vast number of advertisers competing in such auctions and the associated high stakes. Learning how much to bid would help an advertiser to improve his/her revenue. Motivated by this, we investigate on how a rational advertiser would compute the bid prices in a sponsored search auction and can learn optimal bidding strategies.

We now briefly clarify two characteristics - rationality and intelligence - of advertisers. An advertiser is *rational* in the game theoretic sense of making decisions consistently in pursuit of his/her own objectives. Each advertiser's objective is to maximize the expected value of his/her own payoff measured in some utility scale. Note that *selfishness* or *self-interest* is an important implication of rationality. Each advertiser is *intelligent* in the game theoretic sense of knowing everything about the underlying game that a game theorist knows and he/she can make any inferences about the game that a game theorist can make. In particular, each advertiser is *strategic*, that is, takes into account his/her knowledge or expectation of behavior of other agents. He/she is capable of doing the required computations. In this paper, we seek to study the bidding behavior of rational and intelligent advertisers. When there is no ambiguity, we use the word *rational* to mean all both rationality and intelligence.

C. Related Work

The work of [1] investigates the Generalized Second Price (GSP) mechanism for sponsored search auction under *static settings*. The work assumes that the value derived out of a single user-click by an advertiser is publicly known to all the rival advertisers, and then they analyze the underlying static one-shot game of complete information. Garg [2] and Siva Sankar [3] have proposed auction mechanisms superior to the GSP mechanism in their work. Feng and Zhang [4] capture the sponsored search auction setting with a dynamic model and identify a Markov perfect equilibrium bidding strategy for the case of two advertisers. The authors find that in such a dynamic environment, the equilibrium price trajectory follows a cyclical pattern. They study real world behavior of bidders based on experimental data. Asdemir

[5] considers a *two advertisers scenario* and shows that bidding war cycles and static bid patterns frequently observed in sponsored search auctions can result from Markov perfect equilibria. Kitts and Leblanc [6] present a trading agent for sponsored search auctions. The authors formulate it as an integer programming problem and solve it using statistically estimated values of the unknowns. The trading agent basically computes how much to bid on behalf of an advertiser, given the bids of the other advertisers. Kitts and LeBlanc [7] describe the lessons learned in deploying an intelligent trading agent for electronic pay-per-click keyword auctions and an auction simulator for demonstration and testing purposes.

Tesauro and Kephart [8] investigate how adaptive software agents may utilize reinforcement learning [9], [10] techniques such as *Q-learning* to make economic decisions such as setting prices in a competitive market place. Our work does a preliminary investigation of how learning might help advertisers in sponsored search auctions to compute optimal bids.

D. Contributions and Outline of the Paper

This paper investigates the bidding behavior of rational advertisers in sponsored search auctions. Following are the contributions of the paper.

- We first formulate an infinite horizon alternative-move game model of advertiser bidding behavior in sponsored search auctions, assuming the generalized second price auction as the auction mechanism. We call this the *sponsored search auction game*.
- We analyze the two advertiser scenario completely and characterize all the pure strategy and mixed strategy Nash equilibria.
- Next we analyze the effect of *myopic best response bidding* by the advertisers. We show that the bid prices will lead to one of the pure strategy Nash equilibria, if the advertises follow myopic best response bidding.
- We then investigate the use of *Q-learning* [11], [10] as the algorithm for learning optimal bid prices and show that the Q-values converge even if both the advertisers learn simultaneously. Our simulations with different payoff matrices show that the set of pure strategy Nash equilibria of the Q-values matrix is a proper subset of those of the immediate payoff matrix.

The sequence in which we progress in this paper is as follows. In Section II, we develop a game theoretic model of advertiser bidding behavior and use this as a building block in the subsequent sections. We also describe the problem formulation. We describe In Section III, we characterize all the pure strategy and mixed strategy *Nash equilibria* for a two-advertiser scenario. In Section IV, we prove that the bid prices lead to a Nash equilibrium, if the advertisers follow *myopic best response bidding*. In Section V, we apply *Q-learning* in this setting and show that the *Q*-values converge even if both the advertisers learn simultaneously. Section VI summarizes and concludes the paper.

II. THE SPONSORED SEARCH AUCTION GAME

Consider n rational advertisers competing for m positions (advertising slots) in a search engine's result page against a keyword k. Let v_i be advertiser i's per-click valuation associated with the keyword. The valuation could be, for example, the expected profit the advertiser can get from one click-through. We take v_i as advertiser i's type. In this setting, in steady state, we assume advertisers already know their competitors' valuations (perhaps after a sufficiently long learning period). Assume that each position j has a positive expected click-through-rate $\alpha_j > 0$. The *click-through-rate* (CTR) is the the expected number of clicks received per unit time by an Ad in position j. Advertiser i's payoff per unit time from being in position j is equal to $\alpha_j v_i$ minus his/her payment to the search engine. Note that these assumptions imply that the number of times a particular position is clicked does not depend on the ads in this and other positions, and also that an advertiser's value per click does not depend on the position in which its ad is displayed. Without loss of generality, positions are labeled in a descending order: for any j and k such that $j < k$, we have $\alpha_j > \alpha_k$. The notation is provided in Table I.

$N = \{1, 2, \ldots, n\}$	:	set of advertisers
$i \in N$	:	index for advertisers
$M = \{1, 2, \ldots, m\}$	:	set of positions
$j \in M$	:	index for positions
V_i	:	type set of advertiser i
$v_i \in V_i$	:	type of advertiser i
	:	value per click to advertiser i
V_{-i}	:	$V_1 \times \ldots \times V_{i-1} \times V_{i+1} \times \ldots \times V_n$
$v_{-i} \in V_{-i}$	:	set of types of all advertisers other than i
α_j	:	expected number of clicks per unit time (CTR - clickthrough rate) for position j
p_i	:	payment by advertiser i when a searcher clicks on his ad
b_i	:	bid of Advertiser i
π_i	:	payoff function of advertiser i

TABLE I

NOTATION FOR THE MODEL

For simplicity we consider that there are only 2 advertisers. We keep using the phrases advertisers, bidders, agents, players interchangeably. Also we use the terms row player for the player 1 and column player for the player 2.

Note that sponsored search auctions have no closing time and the same keyword may keep occurring again and again. The problem, for any given keyword, is to maximize the payoff of an advertiser by learning how much to bid. i.e. maximize $(v_i - p_i) * \alpha_j$, where p_i is the amount the advertiser is required to pay each time the advertiser's link is clicked by a user. Here v_i is known and p_i and α_j depend on the how much each advertiser has bid. So, the problem of the advertiser is to learn how much to bid, given the bids of the other advertisers, so as to maximize his/her payoff. As already stated, for the sake of simplicity, we consider the

case of 2 advertisers. Assume that the standard practice of *GSP* mechanism is being used by search engines. Then, the payoff function is

$$
\begin{aligned}
\pi_i(b_1, b_2) &= (v_i - b_{-i}) * \alpha_1 & \text{if } b_i > b_{-i} \\
&= v_i * \alpha_2 & \text{otherwise.}
\end{aligned}
\tag{1}
$$

In the above, when $b_1 = b_2$, the tie could be resolved in an appropriate way. We assume in the event of a tie that the first slot is allocated to advertiser 1 and the second slot to advertiser 2. Note that b_{-i} indicates the bid of the bidder other than i.

III. NASH EQUILIBRIA OF THE SPONSORED SEARCH AUCTION GAME

Here and throughout the remainder of the paper, we assume that bids are quantized to integers. We first present two illustrative examples.

Example 1: Consider 2 advertisers with valuations $v_1 = 5$ and $v_2 = 5$. Let $\alpha_1 = 3$ and $\alpha_2 = 2$. This would result in Table II (2-player matrix game) with bids as their actions. By rationality assumption no advertiser bids beyond his valuation.

		2					
		0	1	2	3	4	5
1	0	15,10	10,15	10,15	10,15	10,15	10,15
	1	15,10	12,10	10,12	10,12	10,12	10,12
	2	15,10	12,10	9,10	10,9	10,9	10,9
	3	15,10	12,10	9,10	6,10	10,6	10,6
	4	15,10	12,10	9,10	6,10	3,10	10,3
	5	15,10	12,10	9,10	6,10	3,10	0,10

TABLE II

PAYOFF MATRIX OF ADVERTISERS IN EXAMPLE 1

Example 2: Consider 2 advertisers with valuations $v_1 = 8$ and $v_2 = 5$. Let $\alpha_1 = 3$ and $\alpha_2 = 2$. This would result in Table III (2-player matrix game) with bids as their actions. Again, by rationality assumption no advertiser bids beyond his valuation.

		2					
		0	1	2	3	4	5
1	0	24,10	16,15	16,15	16,15	16,15	16,15
	1	24,10	21,10	16,12	16,12	16,12	16,12
	2	24,10	21,10	18,10	16,9	16,9	16,9
	3	24,10	21,10	18,10	15,10	16,6	16,6
	4	24,10	21,10	18,10	15,10	12,10	16,3
	5	24,10	21,10	18,10	15,10	12,10	9,10
	6	24,10	21,10	18,10	15,10	12,10	9,10
	7	24,10	21,10	18,10	15,10	12,10	9,10
	8	24,10	21,10	18,10	15,10	12,10	9,10

TABLE III

PAYOFF MATRIX OF ADVERTISERS IN EXAMPLE 2

A. Pure Strategy Nash Equilibria

In general the structure of the payoff matrix would be as in Table IV. Note that in each cell of the payoff matrix the first entry gives the row player's payoff and the second entry gives the column player's payoff. The row player moves vertically and the column player moves horizontally.

		2					
	0	1	2	3	4	5	6
0	$v_1\alpha_1$ $v_2\alpha_2$	$v_1\alpha_2$ $v_2\alpha_1$	$v_1\alpha_2$ $v_2\alpha_1$	$v_1\alpha_2$ $v_2\alpha_1$	$v_1\alpha_2$ $v_2\alpha_1$	$v_1\alpha_2$ $v_2\alpha_1$	→
1	$v_1\alpha_1$ $v_2\alpha_2$	$(v_1-1)\alpha_1$ $v_2\alpha_2$	$v_1\alpha_2$ $(v_2-1)\alpha_1$	$v_1\alpha_2$ $(v_2-1)\alpha_1$	$v_1\alpha_2$ $(v_2-1)\alpha_1$	$v_1\alpha_2$ $(v_2-1)\alpha_1$	→
2	$v_1\alpha_1$ $v_2\alpha_2$	$(v_1-1)\alpha_1$ $v_2\alpha_2$	$(v_1-2)\alpha_1$ $v_2\alpha_2$	$v_1\alpha_2$ $(v_2-2)\alpha_1$	$v_1\alpha_2$ $(v_2-2)\alpha_1$	$v_1\alpha_2$ $(v_2-2)\alpha_1$	→
3	$v_1\alpha_1$ $v_2\alpha_2$	$(v_1-1)\alpha_1$ $v_2\alpha_2$	$(v_1-2)\alpha_1$ $v_2\alpha_2$	$(v_1-3)\alpha_1$ $v_2\alpha_2$	$v_1\alpha_2$ $(v_2-3)\alpha_1$	$v_1\alpha_2$ $(v_2-3)\alpha_1$	→
4	$v_1\alpha_1$ $v_2\alpha_2$	$(v_1-1)\alpha_1$ $v_2\alpha_2$	$(v_1-2)\alpha_1$ $v_2\alpha_2$	$(v_1-3)\alpha_1$ $v_2\alpha_2$	$(v_1-4)\alpha_1$ $v_2\alpha_2$	$v_1\alpha_2$ $(v_2-4)\alpha_1$	→
5	↓	↓	↓	↓	↓		

(Row label **1** appears on the left side of the table.)

TABLE IV

PAYOFF MATRIX OF THE ADVERTISERS IN A GENERAL SCENARIO

Observe a thick stair case boundary drawn in Table IV. The arrows indicate that all the utility pairs along the arrow would be the same as the entry just before the start of the arrow. That is, (1) below the stair case boundary, all utility pairs in each column would be the same and (2) above the stair case boundary, all utility pairs in each row would be the same. Note that these table entries and the specific structure observed is because of the payoff values given by equation (1). Keeping this in mind and observing that $(v_i - k)\alpha_1 < v_i\alpha_2$ for some $k \geq 0$, we define below the variables k_1 and k_2.

Each row in Table IV corresponds to a bid of advertiser 1. In each row, each column corresponds to advertiser 2's bid, and each cell gives the corresponding payoff of each of the advertiser. Consider a row corresponding to a bid, say, b_1 of advertiser 1. In that row, observe the payoffs of advertiser 2. The advertiser 2's payoff in all cells, to the left of the stair case boundary would be $v_2\alpha_2$, and to the right of the stair case boundary would be $(v_2 - b_1)\alpha_1$. So, as the advertiser 1's bid increases, beyond some threshold bid, the advertiser 2 being rational tries to bid to the left of the stair case boundary. Call that threshold as k_1. Therefore, as long as the advertiser 1 bids below k_1, the advertiser 2 bids to the right of stair case boundary in the corresponding row and vice versa. Formally we can define k_1 as follows: k_1 is the least integer k such that $v_2\alpha_2 > (v_2 - k)\alpha_1$.

On similar lines, we can define k_2. Each column in Table IV corresponds to a bid of advertiser 2. In each column, each row corresponds to advertiser 1's bid, and each cell gives the corresponding payoff of both the advertisers. Consider a column corresponding to a bid, say, b_2 of advertiser 2. In that row, observe the payoffs of advertiser 1. The advertiser 1's payoff in all cells, above the stair case boundary would be $v_1\alpha_2$, and below the stair case boundary would be $(v_1 - b_2)\alpha_1$. So, as the advertiser 2's bid increases, beyond some threshold bid, the advertiser 1 being rational tries to bid to the left of the stair case boundary. Call that threshold as k_2. Therefore, as long as the advertiser 2 bids below k_2, the advertiser 2 bids to the right of stair case boundary in the corresponding column and vice-versa. Formally we can define k_2 as follows: k_2 is the least integer k such that $v_1\alpha_2 > (v_1 - k)\alpha_1$.

Let b_1 and b_2 be the bids of advertiser 1 and 2 respectively. Now we divide Table IV into 4 regions as follows.

Region 1: $b_1 < k_1$ and $b_2 \geq k_2$
Region 2: $b_1 < k_1$ and $b_2 < k_2$
Region 3: $b_1 \geq k_1$ and $b_2 < k_2$
Region 4: $b_1 \geq k_1$ and $b_2 \geq k_2$

This can be viewed as follows. Divide the plane of Table IV into 4 quadrants. Draw the x-axis through the lines between the rows corresponding to bids $(k_1 - 1)$ and k_1. Similarly, draw the y-axis through the lines between the columns corresponding to bids $(k_2 - 1)$ and k_2. Now the first quadrant corresponds to *Region 1*, the second quadrant corresponds to *Region 2* and so on.

From the above analysis, based on the definitions of k_1 and k_2 and the specific structure of the game shown in Table IV, it is easy to see that if the advertisers start their bidding in any cell of either *Region 1* or *Region 3*, then neither advertiser would move away unilaterally from his/her bid.

However, if they start their bidding in any cell of *Region 2*, then either the row player tries to move down or the column player tries to move right. Similarly, if they start their bidding in any cell of *Region 4*, then either the row player tries to move up or the column player tries to move left.

We state below an interesting proposition which is based on the above facts.

Proposition 1: The bids corresponding to cells in *Region 1* and *Region 3* of Table IV are the only pure strategy Nash equilibria, $(s_1^*(v_1, v_2), s_2^*(v_1, v_2))$. That is,

$$(s_1^*(v_1, v_2), s_2^*(v_1, v_2)) \in$$
$$\{(b_1, b_2) : \quad (0 \leq b_1 < k_1 \text{ and } k_2 \leq b_2 \leq v_2)$$

$$\text{or} \quad (k_1 \leq b_1 \leq v_1 \text{ and } 0 \leq b_2 < k_2)\} \quad (2)$$

All pure strategy Nash equilibria for the Examples 1 and 2 are given in Tables V and VI respectively. In those tables, note that, following the above definitions, the top-right shaded region is *Region 1*. The non-shaded region to the left of *Region 1* is *Region 2*. The bottom-left shaded region is *Region 3*, and the non-shaded region to the right of *Region 3* is *Region 4*. The shaded regions, corresponding to *Regions 1 and 3*, comprise the set of all pure strategy Nash equilibria.

B. Mixed Strategy Nash Equilibria

Now, we turn our attention to the set of all mixed strategy Nash equilibria (MSNE). Recall the following property of MSNE. Consider a two player matrix game, with players A and B, with strategy sets $\{a_1, a_2, \ldots, a_n\}$ and $\{b_1, b_2, \ldots, b_m\}$ respectively. Let (σ_A^*, σ_B^*) be a MSNE of

		2					
		0	1	2	3	4	5
	0	15,10	10,15	10,15	10,15	10,15	10,15
	1	15,10	12,10	10,12	10,12	10,12	10,12
1	2	15,10	12,10	9,10	10,9	10,9	10,9
	3	15,10	12,10	9,10	6,10	10,6	10,6
	4	15,10	12,10	9,10	6,10	3,10	10,3
	5	15,10	12,10	9,10	6,10	3,10	0,10

TABLE V

PURE STRATEGY NASH EQUILIBRIA OF EXAMPLE 1 (SHADED REGIONS)

		2					
		0	1	2	3	4	5
	0	24,10	16,15	16,15	16,15	16,15	16,15
	1	24,10	21,10	16,12	16,12	16,12	16,12
	2	24,10	21,10	18,10	16,9	16,9	16,9
1	3	24,10	21,10	18,10	15,10	16,6	16,6
	4	24,10	21,10	18,10	15,10	12,10	16,3
	5	24,10	21,10	18,10	15,10	12,10	9,10
	6	24,10	21,10	18,10	15,10	12,10	9,10
	7	24,10	21,10	18,10	15,10	12,10	9,10
	8	24,10	21,10	18,10	15,10	12,10	9,10

TABLE VI

PURE STRATEGY NASH EQUILIBRIA OF EXAMPLE 2 (SHADED REGIONS)

the game. $(\sigma_A^*, \sigma_B^*) \in (\Delta(a_1, a_2, \ldots, a_n), \Delta(b_1, b_2, \ldots, b_m))$ is a MSNE of players A and B if and only if both of the following 2 conditions hold.

1) if player A plays σ_A^*, for each of actions $b \in \{b : \sigma_B^*(b) > 0\}$, the player B should get equal and non-negative payoff and, for any action $b \in \{b : \sigma_B^*(b) = 0\}$, the player B should get a payoff less than or equal to the payoff of any action in $\{b_1, b_2, \ldots, b_m\}$

2) similarly, if player B plays σ_B^*, then, for each of actions $a \in \{a : \sigma_A^*(a) > 0\}$, the player A should get equal payoff and, for any action $a \in \{a : \sigma_A^*(a) = 0\}$, the player A should get a payoff less than or equal to the payoff of any action in $\{a_1, a_2, \ldots, a_n\}$

Keeping this property in mind and noting the generic structure of the game shown in Table IV, we make the following observation.

Proposition 2: The mixed strategy profile $(\sigma_1^*(v_1, v_2), \sigma_2^*(v_1, v_2))$ is a mixed strategy Nash equilibrium if and only if

1) σ_1^* is any mixed strategy over advertiser 1's pure strategies in *Region 1* and σ_2^* is any mixed strategy over advertiser 2's pure strategies in *Region 1* **or**

2) σ_1^* is any mixed strategy over advertiser 1's pure strategies in *Region 3* and σ_2^* is any mixed strategy over advertiser 2's pure strategies in *Region 3*

That is, $(\sigma_1^*(v_1, v_2), \sigma_2^*(v_1, v_2)) \in$

$\{(b_1, b_2) : \quad (b_1 \in \Delta(0, \ldots, k_1 - 1) \text{ and } b_2 \in \Delta(k_2, \ldots, v_2))$

$or \ (b_1 \in \Delta(k_1, \ldots, v_1) \text{ and } b_2 \in \Delta(0, \ldots, k_2 - 1))\}$ (3)

To prove the above proposition, we first note that the *if part* is easy to see. To prove the the *only if part*, we first make the following observations.

- The *MSNE* cannot be over *Region 2*
  That is, $(\sigma_1^*(v_1, v_2), \sigma_2^*(v_1, v_2)) \notin \{(b_1, b_2) : \quad b_1 \in \Delta(0, \ldots, k_1 - 1) \text{ and } b_2 \in \Delta(0, \ldots, k_2 - 1)\}$
- The *MSNE* cannot be over *Region 4*
  That is, $(\sigma_1^*(v_1, v_2), \sigma_2^*(v_1, v_2)) \notin \{(b_1, b_2) : \quad b_1 \in \Delta(k_1, \ldots, v_1) \text{ and } b_2 \in b_2 \in \Delta(k_2, \ldots, v_2)\}$
- $\sigma_1^*(v_1, v_2)$ cannot have positive probabilities for pure strategies in both $\{0, \ldots, k_1 - 1\}$ and $\{k_1, \ldots, v_1\}$.
  This can be seen through a simple contradiction.
- $\sigma_2^*(v_1, v_2)$ cannot have positive probabilities for pure strategies in both $\{0, \ldots, k_2 - 1\}$ and $\{k_2, \ldots, v_1\}$.
  This too can be seen through a simple contradiction.

The above proposition and proof can be better understood and can be easily verified through the Examples 1 and 2 in Tables V and VI respectively.

IV. ANALYSIS OF MYOPIC BEST RESPONSE BIDDING BY THE ADVERTISERS

A myopic optimal (*myoptimal* for short) bidding policy of advertiser i is obtained by choosing the bid b_i that maximizes $\pi_i(b_i, b_{-i})$. This can be represented as a response function $R_i(b_{-i})$:

$$R_i(b_{-i}) = arg \max_{b_i} \pi_i(b_i, b_{-i}) \quad (4)$$

Starting from any given initial bid vector, one can successively apply the response functions R_1 and R_2. This would lead to an equilibrium bid vector, from where neither advertiser would like to move away unilaterally. This can be proved as follows.

Based on the definitions of k_1 and k_2 and the specific structure of the game shown in Table IV, it is easy to see that

1) If the advertisers start their bidding in any cell of either *Region 1* or *Region 3*, then neither advertiser would move away from his/her bid unilaterally. Hence, that itself corresponds to an equilibrium bid vector.

2) If the two advertisers start their bidding in any cell of *Region 2*, then either the row player tries to move down or the column player tries to move right. In a finite number of steps, this would lead to both the players entering either *Region 1* or *Region 3*, from where neither player would move away from higher bid unilaterally. Hence, that corresponds to an equilibrium bid vector.

3) Similarly, if the two advertisers start their bidding in any cell of *Region 4*, then either the row player tries to move up or the column player tries to move left. Again, in a finite number of steps, this would lead to both the players entering either *Region 1* or *Region 3*, from where neither player would move away from

his/her bid unilaterally. Hence, that corresponds to an equilibrium bid vector.

Based on the above observations, we have the following proposition.

Proposition 3: If both the advertisers follow *myopic best response*, then that would lead to a pure strategy Nash equilibrium in a finite number of steps, no matter what the initial bid vector of the advertisers is.

The above proposition can be better understood and can be easily verified through the Examples 1 and 2 in Tables V and VI respectively. A sample trajectory of bid vector starting from bidding vector (5,5) in Example 1, when both the advertisers follow myopic best response, is shown in Figure 2.
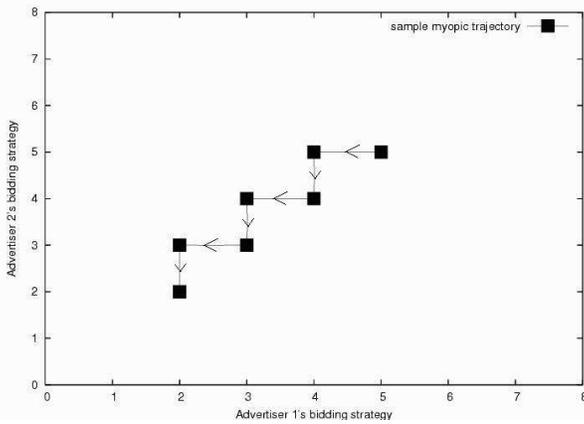

Fig. 2. Sample trajectory of bid-vectors in Example 1 if both advertisers follow myopic

V. ANALYSIS OF Q-LEARNING BASED BIDDING

We see whether *Q-learning* [11], [10] by one or both of the advertisers raises both of their profits substantially from what they would obtain using myopic best-response bidding. There are two fundamentally different approaches of applying *Q-learning* in this multi-agent setting.

- The first approach is to construct the long-term payoff matrix (by learning *Q-values* using immediate payoffs such as in Tables 2, 3 and 4), to see whether the Nash equilibria of that payoff matrix are a strict subset of the Nash equilibria of the original (immediate) payoff matrix, to see whether a myopic policy on that payoff matrix would yield a better payoff, etc. To understand long-term payoff, consider player 1 playing action a_1. Then if player 2 follows, say, b_2 then it may give him the maximum immediate payoff. But then, player 1 can now take a different counter move which may worsen the payoff of player 2. So, the long-term payoff matrix, in some appropriate sense, gives the payoffs that summarize the other player's counter-moves. We pursue this approach in our current work.
- The second approach is to design a *multi-agent Q-learning algorithm* that leads to a mixed strategy that would improve payoff for both advertisers.

How can the advertisers introduce foresight into their pricing decisions? One way is to optimize the cumulative future discounted profit instead of immediate profit. One of the advertisers i adds expected profit from the current move to future discounted profit. Projecting further into the future, advertiser i computes the expected future discounted profit as an infinite weighted sum of expected profits in future time steps. More compactly, we define

$$Q_i(b_i, b_{-i}) = \pi_i(b_i, b_{-i}) + \beta \max_{b_i} Q_i(b_i, R_{-i}(b_i)) \quad (5)$$

where β is a discount parameter, ranging from 0 to 1, and R_{-i} represents the policy that optimizes Q_{-i}:

$$R_{-i}(b_i) = arg \max_{b_{-i}} Q_{-i}(b_{-i}, b_i) \quad (6)$$

The Q-values are obtained by adding immediate profit of taking action (bid) b_i to the best Q-value (long term profit) that could be obtained (by taking into consideration, the opponent's response action R_{-i}). This formulation is a straight forward extension of single agent Markov decision process. It is well established that, if R_{-i} represents any fixed policy (myoptimal or otherwise), then a reasonable updating procedure will yield a unique, stable future discounted profit landscape $Q_i(b_i, b_{-i})$ and associated response function $R_i(b_{-i})$. However, if the other advertiser similarly computes Q_{-i} and the associated R_{-i}, then a unique fixed point is not guaranteed.

However, in the current setting, we found experimentally that in both the cases that the Q-values converge. This was found through a standard Q-learning updating procedure in which, starting from initial functions Q_i and Q_{-i}, a random seller and a random price vector were chosen, the right hand side of Equation (5) was evaluated for that advertiser and the bid vector using Equation (6), and the Q-value for that advertiser and price vector were moved toward this computed value by a fraction β that diminished gradually with time.

Table VII shows the converged Q-values for single advertiser learning, with the other advertiser following myopic best response, for Example 1. Table VIII shows the converged Q-values when both the advertisers simultaneous learn in Example 1. These tables are obtained through simulations. Observe that the set of pure strategy Nash equilibria of these Q-value matrices is a proper subset of those of the immediate payoff matrix given in Table V.

VI. CONCLUSIONS AND FUTURE WORK

A. Conclusions

In this paper, we developed an infinite horizon alternative-move game of advertiser bidding behavior. We analyzed the two advertiser scenario and we characterized all the Nash equilibria. Next, we showed that the bid prices lead to one of the pure strategy Nash equilibria, if the advertises follow myopic best response bidding. We then applied Q-learning in this setting and found that the Q-values converge even if both the advertisers learn simultaneously. We conducted extensive simulations with different payoff matrices and found that, in

		2					
		0	1	2	3	4	5
1	0	18.1500, 13.7499	13.1499, 18.1500	13.1499, 18.7499	13.1499, 18.7499	13.1499, 18.7499	13.1499, 18.7499
	1	17.6299, 13.7499	14.6299, 13.1499	12.6299, 15.7499	12.6299, 15.7499	12.6299, 15.7499	12.6299, 15.7499
	2	18.7499, 13.7499	15.7499, 13.1499	12.7499, 13.7499	13.7499, 12.7499	13.7499, 12.7499	13.7499, 12.7499
	3	18.7499, 13.7499	15.7499, 13.1499	12.7499, 13.7499	9.7499, 13.7499	13.7499, 9.7499	13.7499, 9.7499
	4	18.7499, 13.7499	15.7499, 13.1499	12.7499, 13.7499	9.7499, 13.7499	6.7499, 13.7499	13.7499, 6.7499
	5	18.7499, 13.7499	15.7499, 13.1499	12.7499, 13.7499	9.7499, 13.7499	6.7499, 13.7499	3.7499, 13.7499

TABLE VII

CONVERGED Q-VALUES OF EACH ADVERTISER, FOR EXAMPLE 1, WHEN THE OTHER ADVERTISER FOLLOWS *myopic best response*

		2					
		0	1	2	3	4	5
1	0	17.4999, 12.7500	12.4999, 17.7500	12.4999, 18.7499	12.4999, 18.7499	12.4999, 18.7499	12.4999, 18.7499
	1	17.4999, 12.7500	14.4999, 12.7500	12.4999, 15.7499	12.4999, 15.7499	12.4999, 15.7499	12.4999, 15.7499
	2	17.4999, 12.7500	14.4999, 12.7500	11.4999, 13.7499	12.4999, 12.7499	12.4999, 12.7499	12.4999, 12.7499
	3	17.4999, 12.7500	14.4999, 12.7500	11.4999, 13.7499	8.4999, 13.7499	12.4999, 9.7499	12.4999, 9.7499
	4	17.4999, 12.7500	14.4999, 12.7500	11.4999, 13.7499	8.4999, 13.7499	5.4999, 13.7499	12.4999, 6.7499
	5	17.5000, 12.7500	14.5000, 12.7500	11.5000, 13.7499	8.5000, 13.7499	5.5000, 13.7499	2.5000, 13.7499

TABLE VIII

CONVERGED Q-VALUES IN THE CASE OF *simultaneous Q-learning*, FOR EXAMPLE 1.

each case, the set of pure strategy Nash equilibria of the Q-values matrix is a proper subset of those of the immediate payoff matrix.

B. Future Work

This paper opens up plenty of room for further investigation:

- Investigating how the single-agent *Q-learning* and simultaneous-*Q-learning* would affect the expected payoff of the advertiser, using the long term payoff matrices (Q-values) that we obtained through simulations.
- Designing a *multi-agent Q-learning algorithm* [12] that leads to a mixed strategy which would improve payoff for both advertisers.
- Extending the findings of this study to the case of multiple (greater than 2) advertisers
- Incorporating *budget constraints* that the advertisers may have

REFERENCES

[1] B. Edelman and M. Ostrovsky. Strategic bidder behavior in sponsored search auctions. In *Workshop on Sponsored Search Auctions in conjunction with the ACM Conference on Electronic Commerce*, 2005.
[2] Dinesh Garg. *Design of Innovative Mechanisms for Contemporary Game Theoretic Problems in Electronic Commerce*. PhD thesis, Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India, May 2006.
[3] S. Siva Sankar Reddy. Designing an optimal auction and learning bid prices in sponsored search auctions. Technical report, Master Of Engineering Dissertation, Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India, May 2006.
[4] J. Feng and X. Zhang. Price cycles in online advertising auctions. In *International Conference on Information Systems*, 2005.
[5] K. Asdemir. Bidding patterns in search engine auctions. Working Paper, Department of Accounting and Management Information Systems, University of Alberta School of Business, 2004.
[6] B. Kitts and B. Leblanc. Optimal bidding on keyword auctions. *Electronic Markets*, 14(3):186–201, 2004.
[7] B. Kitts and B. LeBlanc. A trading agent and simulator for keyword auctions. In *International Joint Conference on Autonomous Agents and Multiagent Systems*, 2004.
[8] G. Tesauro and J.O. Kephart. Pricing in agent economies using multi-agent Q-learning. In *Workshop on Decision Theoretic and Game Theoretic Agents*, 1999.
[9] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.
[10] D. P. Bertsekas and J. Tsitsiklis. *Neuro-dynamic Programming*. Athena Scientific, Boston, MA, USA, 1996.
[11] C. J. C. H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8:279–292, 1992.
[12] M. L. Littman. Friend-or-foe q-learning in general-sum games. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 322–328, 2001.