Northwestern University
2001 Sheridan Road 580 Leverone Hall Evanston, IL 60208-2014 USA

Discussion Paper #1450
July 13, 2007

"Dynamic Cost-Per-Action Mechanisms and Applications to Online Advertising"

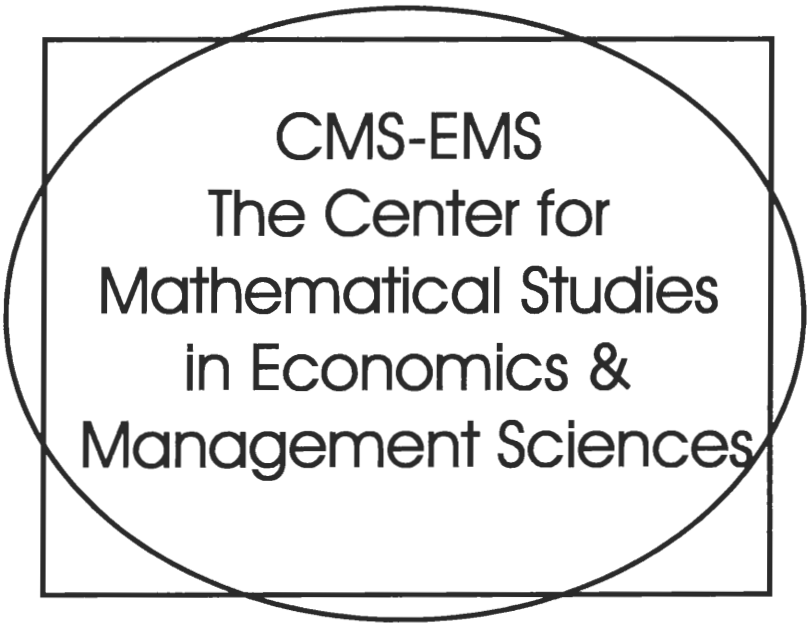
JEL code: D44
Key words: dynamic auctions, sponsored search

Hamid Nazerzadeh
Stanford University

Amin Saberi
Stanford University

Rakesh Vohra
Northwestern University

www.kellogg.northwestern.edu/research/math



Dynamic Cost-Per-Action Mechanisms and Applications to Online Advertising

Hamid Nazerzadeh* Amin Saberi† Rakesh Vohra‡

July 13, 2007

Abstract

We examine the problem of allocating a resource repeatedly over time amongst a set of agents. The utility that each agent derives from consumption of the item is private information to that agent and, prior to consumption may be unknown to that agent. The problem is motivated by keyword auctions, where the resource to be allocated is a slot on a search page. We describe a mechanism based on a sampling-based learning algorithm that under suitable assumptions is asymptotically individually rational, asymptotically Bayesian incentive compatible and asymptotically ex-ante efficient. The mechanism can be interpreted as a cost per action keyword auction.

1 Introduction

We study the following problem: there are a number of self-interested agents competing for identical items sold repeatedly at times $t = 1, 2, \cdots$. At each time t, a mechanism allocates the item to one of the agents. Agents *discover* their utility for the good only if it is allocated to them. If agent i receives the good at time t, she realizes utility u_{it} (denominated in money) for and reports (not necessarily truthfully) the realized utility to the mechanism. Then, the mechanism determines how much the agent has to pay for receiving the item. We allow the utility of an agent to change over time.

For this environment we are interested in auction mechanisms which have the following four properties.

1. The mechanism is individually rational in each period.
2. Agents have an incentive to truthfully report their realized utilities.
3. The efficiency (and revenue) is, in an appropriate sense, not too small compared to a second price auction.

*Management Science and Engineering Department, Stanford University, CA 94305, email: {hamidnz@stanford.edu}

†Management Science and Engineering Department, Stanford University, CA 94305, email: {saberi@stanford.edu}

‡Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, IL 60208, email: {r-vohra@kellogg.northwestern.edu}

4. The correctness of the mechanism does not depend on an a-priori knowledge of the distribution of the u_{it}'s. This feature is motivated by the Wilson doctrine ([21]).

The precise manner in which these properties are formalized is described in section 2.

Each mechanism in the class we investigate is associated with a sampling-based learning algorithm. The learning algorithm is used to estimate the expected utility of the agents, and consists of two alternating phases: explore and exploitation. During an explore phase, the item is allocated for free to a randomly chosen agent. During an exploitation phase, the mechanism allocates the item to the agent with the highest estimated expected utility. After each allocation, the agent who has received the item reports its realized utility. Subsequently, the mechanism updates the estimate of utilities and determines the payment.

Since there is uncertainty about the utilities, it is possible that in some periods the item is allocated to an agent who does not have the highest utility in that period. Hence, the natural second-highest price payment rule would violates individual rationality. If the mechanism does not charge an agent because her reported utility after the allocation is low, it gives her an incentive to shade her reported utility down. Our mechanism solves these problems by using an adaptive, cumulative pricing scheme.

We give sufficient conditions on the underlying learning algorithm that ensure that the corresponding mechanism has the four desired properties. In particular, we identify simple mechanisms that have the desired properties for the case when the u_{it}'s are independent and identically-distributed random variables or where their expected values evolve like independent reflected Brownian motions. In these cases the mechanism is actually *ex-post* individually rational.

In the next section, we will motivate our work in the context of online advertising. However, the motivation for our mechanism is not limited to such applications.

1.1 Keyword auctions

Each keyword query in a search engine returns a page containing links relevant to the query and an ordered list of paid advertisements called sponsored links. For instance, if a seller of novelty gifts buys the word *gift*, each time a user performs a search on this word, a link to the novelty seller will appear on the search results page. When the user clicks on that link, she is sent to the relevant advertiser's web page. The advertiser then pays the search engine for sending the user to its web page. This is called 'cost-per-click' (CPC) pricing.

The price an advertiser pays is determined by a generalized second price (GSP) auction. Advertisers submit bids that represent the CPC they are prepared to pay. The highest bidder wins the top slot, the second highest bidder the second slot and so on. The advertiser in position n pays a price per click one cent higher than the bid of the advertiser in position $(n+1)$.[1]

CPC is considered more attractive than the cost-per-impression (CPM) charging scheme used in traditional media (e.g., magazines and television) or banner advertising. In CPM an advertiser is charged based on the (estimated) number of people exposed to the ad.

CPC's chief drawback is its vulnerability to click-fraud. Click fraud refers to clicks generated by someone or something with no genuine interest in the advertisement. Such clicks can be generated by the publisher of the content who has an interest in receiving a share of the revenue of the advertisement or by a rival who wishes to increase the cost of advertising for the advertiser.

[1] The payments are adjusted by fudge factors that account for the relevance of ads.

Click-fraud is considered by many experts to be the biggest challenge facing the online advertising industry[12, 8, 20, 18].

A natural solution for the problem of click fraud is to charge advertisers according to Cost-Per-Action or Cost-Per-Acquisition (CPA). Instead of paying per click, the advertiser pays only when a user takes a specific action (eg downloads software) or completes a transaction. The relevant set of actions is chosen by the advertiser in advance. Several Companies like Advertising.com, Turn.com, and Snap.com sell advertising in this way. Google and eBay have begun to sell some of their advertising space via CPA [2].

If an action is defined as a sale, then CPA makes generating a fraudulent action a more costly enterprise, but not impossible. One could use a stolen credit card number for example. On the flip side, CPA increases the incentives for the advertiser to under report the number of actions that have taken place so as to reduce her payments. This can be mitigated, but not entirely, by requiring the advertiser to install software that will monitor actions that take place on their web site. Still, even moderately sophisticated advertisers can find a way to manipulate the software if they find it sufficiently profitable.

The main difference between an auction where payments depend on CPC and one where they depend on CPA is that a click can be observed by both the advertiser and the search engine but only the advertiser can observe the action of the user and she can hide it from the search engine. It is this difference that motivates the present paper.

In our setting, the item being allocated is a search query for a keyword. An advertiser obtains a payoff when the user clicks on her advertisement and takes a specific action. Since the payoff is uncertain, she cannot know what it will be unless her ad is displayed. For simplicity of exposition only, we assume one keyword and one advertisement slot. In section 6 we outline how to extend our results to the case where more than one advertisement can be displayed for each query.

1.2 Related Work

Here we summarize some of the mostly closest results from dynamic mechanism design literature (for a comprehensive survey see [19]).

First, in a finitely repeated version of the environment considered here, Athey and Segal [2] construct an efficient, budget balanced, direct revelation mechanism where truthful revelation in each period is Bayesian incentive compatible. The mechanism for multiple periods is obtained by backward induction and an iterative re-balancing of the payments, to achieve a budget balance. Bapna and Weber [4] consider the infinite horizon version of [2]. They describe a class of mechanisms based on the Gittins index (see [9]) and give necessary and sufficient conditions for such mechanisms to be incentive compatible. Bergemann and Välimäki [5] propose an incentive compatible generalization of the Vickrey-Clark-Groves mechanism based on the marginal contribution of each agent for this environment. Recently, Cavallo et. al [7] extend the result when the size of the population changes over time. All these mechanisms need the exact solution of the underlying optimization problems, and therefore require complete information about the prior of the utilities of the agents; also, they do not apply when the evolution of the utilities of the agents is not stationary over time. This violates the last of our desiderata.

In the context of sponsored search, attention has focused on ways of estimating click through

[2]CPA auctions have other advantages like simplifying the bidding language or lowering the barrier to entry for advertisers. For a detailed discussion of the pluses and minuses of CPA auctions see [16].

rates. The obvious way to estimate click through rates would be to sample by assigning an advertiser to a slot independent of their bid just so as to collect data. This reduces revenue and can encourage bidders to shade their bids down. Gonen and Pavlov [10] give a mechanism which learns the click-through rates via optimal sampling and show that truthful bidding is, with high probability, a (weakly) dominant strategy in this mechanism. In [14, 11] and [13] the vulnerability of various procedures for estimating click through rates is examined. Immorlica et. al. [13], in particular, identify a class of click through learning algorithms in which fraudulent clicks can not increase the expected payment per impression by more than $o(1)$. In all of these papers, unlike ours, the utilities of agents are assumed fixed over time.

2 Definitions and Notation

Suppose n agents competing in each period for a single item. The item is sold repeatedly at time $t = 1, 2, \cdots$. Denote by u_{it} the nonnegative utility of agent i for the item at time t. Utilities are denominated in a common monetary scale.

The utilities of agents may evolve over time according to a stochastic process. We assume that for $i \neq j$, the evolution of u_{it} and u_{jt} are independent stochastic processes. We also define $\mu_{it} = E[u_{it}|u_{i1}, \cdots, u_{i,t-1}]$. Throughout this paper, expectations are taken conditioned on the complete history. For simplicity of notation, we now omit those terms that denote such a conditioning. With notational convention, it follows, for example, that $E[u_{it}] = E[\mu_{it}]$. Here the second expectation is taken over all possible histories.

Let $\mathcal{M}$ be a mechanism used to sell the items. At each time, $\mathcal{M}$ allocates the item to one of the agents. Let i be the agent who has received the item at time t. Define x_{it} to be the indicator variable of allocation of the item to i at time t. After the allocation, agent i observes her utility, u_{it}, and then reports r_{it}, as her utility for the item, to the mechanism. Note that we do not require an agent know its utility for possessing the item in advance of acquiring it. The mechanism then determines the payment, denoted by p_{it}.

Definition 1 *An agent i is* truthful *if $r_{it} = u_{it}$, for all time $x_{it} = 1, t > 0$.*

Our goal is to design a mechanism which has the following properties:

Individual Rationality: A mechanism is *ex-post* individually rational if for any time $T > 0$ and any agent $1 \leq i \leq n$, the total payment of agent i does not exceed the sum of her reports:

$$\sum_{t=1}^{T} x_{it} r_{it} - p_{it} > 0.$$

M is *asymptotically ex-ante* individually rational if:

$$\liminf_{T \to \infty} E[\sum_{t=1}^{T} x_{it}\mu_{it} - p_{it}] \geq 0.$$

Asymptotic Incentive Compatibility: This property implies that truthfulness defines an asymptotic Bayesian Nash equilibrium. Consider agent i and suppose all agents except i are truthful.

Let $U_i(T)$ be the expected total profit of agent i, if agent i is truthful between time 1 and T. Also, let $\widetilde{U}_i(T)$ be the maximum of expected profit of agent i under any other strategy. *Asymptotic incentive compatibility* requires that

$$\widetilde{U}_i(T) - U_i(T) = o(U_i(T)).$$

Ex-ante Efficient: Call a mechanism that allocates at each time t (and for each history) the item to an agent in $\arg\max_i \mu_{it}$ ex-ante efficient. If each agent i knew μ_{it} for all t, such an allocation could be achieved in an an incentive compatible way. Have each i report a value of μ_{it}, give the item to any agent in $\arg\max_i \mu_{it}$ and charge them the second highest μ_{it}. Let γ_t be the second highest μ_{it} at time $t > 0$. Then, the expected revenue of this second price mechanism is equal to $E[\sum_{t=1}^{T} \gamma_t]$.

We will measure how close $\mathcal{M}$ comes to being ex-ante efficient by comparing its expected revenue the expected revenue of the second price mechanism just described. Let $R(T)$ be the expected revenue of mechanism $\mathcal{M}$ between time 1 and T when the agents are truthful, i.e. $R(T) = E[\sum_{t=1}^{T} \sum_{i=1}^{n} p_{it}]$.

Then, $\mathcal{M}$ is *ex-ante efficient* if

$$E[\sum_{t=1}^{T} \gamma_t] - R(T) = o(R(T)).$$

3 Proposed Mechanism

The mechanism we propose is built around a learning algorithm that estimates the expected utility of the agents. We refrain from an explicit description of the learning algorithm. Rather, we describe sufficient conditions for the learning algorithm that ensure that the resulting mechanism has the three properties we seek (see section 3.1). In section 4 and 5 we give two examples of environments where learning algorithms satisfying these sufficient conditions exist.

The mechanism consists of two phases: *explore* and *exploit*. During the explore phase, with probability $\eta(t)$, $\eta : \mathbb{N} \to [0, 1]$, the item is allocated for free to a randomly chosen agent. During the exploit phase, the mechanism allocates the item to the agent with the highest estimated expected utility. Afterwards, the agent reports her utility to the mechanism and the mechanism determines the payment. The mechanism is give in Figure 1. We first formalize our assumptions about the learning algorithm and then we discuss the payment scheme.

The learning algorithm, samples u_{it}'s at rate $\eta(t)$ and based on the history of the reports of agent i, returns an estimate of μ_{it}. Let $\widehat{\mu}_{it}(T)$ be the estimate of the algorithm for μ_{it} conditional on the history of the reports up to time T. The history of the reports of agent i up to time T is the sequence of the reported values and times of observation of u_{it} up to but not including time T. Note that we allow $T > t$. Thus, information at time $T > t$ can be used to revise an estimate of μ_{it} made at some earlier time. We assume that increasing the number of samples only increases the accuracy of the estimations, i.e. for any truthful agent i, and times $T_1 \leq T_2$:

$$E[|\widehat{\mu}_{it}(T_1) - \mu_{it}|] \quad \geq E[|\widehat{\mu}_{it}(T_2) - \mu_{it}|]. \tag{1}$$

In the inequality above, and in the rest of the paper, the expectations of $\widehat{\mu}_{it}$ are taken over the evolution of u_{it}'s and the random choices of the mechanism. Recall that for simplicity of notation, we omit the terms denoting conditional expectations.

For $t = 1, 2, \ldots$

With probability $\eta(t)$, *explore*:

Randomly allocate the item to an agent i, $1 \leq i \leq n$.

$p_{it} \leftarrow 0$

With probability $1 - \eta(t)$, *exploit*:

Randomly allocate the item to an agent $i \in \text{argmax}_i\{\widehat{\mu}_{it}(t)\}$.

$p_{it} \leftarrow \sum_{k=1}^{t-1} y_{ik} \min\{\widehat{\gamma}_k(t), \widehat{\mu}_{ik}(k)\} - \sum_{k=1}^{t-1} p_{ik}$

$r_{it} \leftarrow$ the report of agent i.

$p_{jt} \leftarrow 0$, $j \neq i$

Figure 1: Mechanism $\mathcal{M}$

To describe the payments recall that γ_t is the second highest μ_{it} and let $\widehat{\gamma}_t(T) = \max_{j \neq i}\{\widehat{\mu}_{jt}(T)\}$, where i is the agent who received the item at time t. We define y_{it} to be the indicator variable of the allocation of the item to agent i during an exploit phase. The payment of agent i at time t, denoted p_{it}, is determined so that:

$$\sum_{k=1}^{t} p_{ik} = \sum_{k=1}^{t-1} y_{ik} \min\{\widehat{\gamma}_k(t), \widehat{\mu}_{ik}(k)\}.$$

An agent only pays for items that are allocated to him during the exploit phase, up to but not including time t. At time t, the payment of agent i for the item she received at time $k < t$ is $\min\{\widehat{\gamma}_k(t), \widehat{\mu}_{ik}(k)\}$. Since the learning algorithm's estimates of the utility of the agent become more precise over time, our cumulative payment scheme allows one to correct for errors in the past. Furthermore, because of the estimation errors, it is possible that the mechanism allocates an item to an agent who is not the one with the highest expected utility. By taking the minimum of $\widehat{\gamma}_k(t)$ and $\widehat{\mu}_{ik}(k)$ we avoid overcharging an agent and therefore remove any incentives for an agent to shade down her reported utility.

3.1 Sufficient Conditions

We start with a condition that guarantees asymptotic ex-ante individual rationality and asymptotic incentive compatibility. Let $\Delta_t = \max_i\{|\widehat{\mu}_{it}(t) - \mu_{it}|\}$.

Theorem 1 *If for the learning algorithm:*

$$(C1) \qquad E[\mu_{iT} + \sum_{t=1}^{T-1} \Delta_t] = o(E[\sum_{t=1}^{T} \eta(t)\mu_{it}]), \quad \forall\ 1 \leq i \leq n$$

then mechanism $\mathcal{M}$ is asymptotically ex-ante individually rational and incentive compatible.

We outline the proof first. As we prove in Lemma 2, by condition $(C1)$, the expected profit of a truthful agent up to time T is $\Omega(E[\sum_{t=1}^{T}\eta(t)\mu_{it}])$. This implies asymptotic individual rationality. Also, observe that the expected total error in the estimates of the payments up to time $T-1$ is bounded by $O(E[\sum_{t=1}^{T-1}\Delta_t])$. A non-truthful agent may exploit the gap between the estimated and the actual expected utilities to increase her profit. However, this profit cannot exceed the total error. Also, if i received the item at time t, she would not pay for it before the next time she gets the item during the exploit phase. The utility of this item, contributes tohers profit up to time T. Condition $(C1)$ implies that the expected profit of an agent cannot be increased by more than a $o(1)$ factor.

Lemma 2 *If condition* $(C1)$ *holds, then the expected profit of a truthful agent* i *up to time* T *under* $\mathcal{M}$*,* $U_i(T)$*, is at least:*

$$(\frac{1}{n}-o(1))E[\sum_{t=1}^{T}\eta(t)\mu_{it}].$$

Proof : The items that agent i receives during the explore phase are free. The expected total utility of i for these items up to time T is $\frac{1}{n}E[\sum_{t=1}^{T}\eta(t)\mu_{it}]$. Let $C_T = \{t < T | y_{it} = 1, \text{ if } i \text{ is truthful}\}$ be the subset of periods within an exploit phase.

$$\begin{aligned} U_i(T) &= E[\sum_{t=1}^{T} x_{it}u_{it} - p_{it}] \\ &= E[\sum_{t\notin C_T} x_{it}u_{it}] + E[\sum_{t\in C_T}(u_{it}-p_{it})] \\ &= \frac{1}{n}E[\sum_{t=1}^{T}\eta(t)\mu_{it}] + E[\sum_{t\in C_T}(\mu_{it}-\min\{\widehat{\gamma}_t(T),\widehat{\mu}_{it}(t)\})] \end{aligned} \tag{2}$$

For $t \in C_T$:

$$\begin{aligned} E[(\mu_{it}-\min\{\widehat{\gamma}_t(T),\widehat{\mu}_{it}(t)\})I(t\in C_T)] &\geq E[(\mu_{it}-\widehat{\mu}_{it}(t))I(t\in C_T)] \\ &\geq -E[|\mu_{it}-\widehat{\mu}_{it}(t)|] \\ &\geq -E[\Delta_t] \end{aligned}$$

Substituting into inequality (2), by condition $(C1)$:

$$U_i(T) \geq \frac{1}{n}E[\sum_{t=1}^{T}\eta(t)\mu_{it}] - E[\sum_{t=1}^{T-1}\Delta_t] = \frac{1}{n}E[\sum_{t=1}^{T}\eta(t)\mu_{it}] - o(E[\sum_{t=1}^{T}\eta(t)\mu_{it}]) \tag{3}$$

□

Proof of Theorem 1: Lemma 2 yields asymptotic ex-ante individual rationality. We show that truthfulness is asymptotically a best response when all other agents are truthful. Fix an agent i intending to deviate and let $\mathcal{S}$ be the strategy she deviates to. Fixing the evolution of all u_{jt}'s, $1 \leq j \leq n$, and all random choices of the mechanism, i.e. the steps in the explore phase and the randomly chosen agents, let D_T be the times that i receives the item under strategy

$\mathcal{S}$ during the exploit phase, i.e. $D_T = \{t < T | y_{it} = 1, \text{ if the strategy of } i \text{ is } \mathcal{S}\}$. Similarly, let $C_T = \{t < T | y_{it} = 1, \text{ if } i \text{ is truthful}\}$ be the set of times that i would receive the item during the exploit phase if she is truthful up to T. Also, let $\widehat{\mu}'_{it}$, and $\widehat{\gamma}'_t$ correspond to the estimates of the mechanism when the strategy of i is $\mathcal{S}$. We first compute the expected profit of i, under strategy $\mathcal{S}$, during the exploit phase:

$$\begin{aligned} E[\sum_{t=1}^{T} y_{it}u_{it} - p_{it}] &= E[\sum_{t\in D_T} \mu_{it} - \min\{\widehat{\gamma}'_t(T), \widehat{\mu}'_{it}(t)\}] + E[y_{iT}\mu_{iT}] \\ &= E[\sum_{t\in D_T\setminus C_T} \mu_{it} - \min\{\widehat{\gamma}'_t(T), \widehat{\mu}'_{it}(t)\}] + \\ &\quad E[\sum_{t\in D_T\cap C_T} \mu_{it} - \min\{\widehat{\gamma}'_t(T), \widehat{\mu}'_{it}(t)\}] + \\ &\quad E[y_{iT}\mu_{iT}] \end{aligned} \tag{4}$$

For time $t \geq 1$, we examine two cases:

1. If $t \in D_T \cap C_T$, then agent i, in expectation, cannot decrease the "current price", $\min\{\widehat{\gamma}'_t(T), \widehat{\mu}'_{it}(t)\}$, by more than $O(\Delta_t)$:

$$\begin{aligned} \min\{\widehat{\gamma}'_t(T), \widehat{\mu}'_{it}(t)\} &\geq \min\{\widehat{\gamma}'_t(T), \widehat{\gamma}'_t(t)\} \\ &\geq \gamma_t - \max\{\gamma_t - \widehat{\gamma}'_t(T), \gamma_t - \widehat{\gamma}'_t(t)\} \\ &\geq \gamma_t - (\gamma_t - \widehat{\gamma}'_t(T))^+ - (\gamma_t - \widehat{\gamma}'_t(t))^+ \end{aligned}$$

where $(z)^+ = \max\{z, 0\}$. Recall that $\widehat{\gamma}'_t(T) = \max_{j\neq i}\{\widehat{\mu}'_{it}(T)\}$ and all other agent are truthful. Hence, taking expectation from both sides, by (1):

$$\begin{aligned} E[\min\{\widehat{\gamma}'_t(T), \widehat{\mu}'_{it}(t)\}I(t \in D_T \cap C_T)] &\geq E[(\gamma_t - (\gamma_t - \widehat{\gamma}'_t(T))^+ - (\gamma_t - \widehat{\gamma}'_t(t))^+)I(t \in D_T \cap C_T)] \\ &\geq E[\gamma_t I(t \in D_T \cap C_T)] - E[2\Delta_t] \end{aligned} \tag{5}$$

2. If $t \in D_T \setminus C_T$, agent i cannot increase her "expected profit", $\mu_{it} - \min\{\widehat{\gamma}'_t(T), \widehat{\mu}'_{it}(t)\}$, by more than $O(\Delta_t)$:

$$\begin{aligned} \mu_{it} - \min\{\widehat{\gamma}'_t(T), \widehat{\mu}'_{it}(t)\} &\leq \mu_{it} - \min\{\widehat{\gamma}'_t(T), \widehat{\gamma}'_t(t)\} \\ &\leq (\mu_{it} - \widehat{\mu}_{it}(t)) + (\widehat{\mu}_{it}(t) - \gamma_t) + \max\{\gamma_t - \widehat{\gamma}'_t(T), \gamma_t - \widehat{\gamma}'_t(t)\} \\ &\leq 2\Delta_t + (\gamma_t - \widehat{\gamma}'_t(T))^+ + (\gamma_t - \widehat{\gamma}'_t(t))^+ \end{aligned}$$

Taking expectation from both sides, by (1):

$$\begin{aligned} E[(\mu_{it} - \min\{\widehat{\gamma}'_t(T), \widehat{\mu}'_{it}(t)\})I(t \in D_T - C_T)] &\leq E[2\Delta_t I(t \in D_T - C_T)] \\ &\quad + E[((\gamma_t - \widehat{\gamma}'_t(T))^+ + (\gamma_t - \widehat{\gamma}'_t(t))^+)I(t \in D_T - C_T)] \\ &\leq E[4\Delta_t] \end{aligned} \tag{6}$$

Substituting inequalities (5) and (6) into (4):

$$\begin{aligned} E[\sum_{t=1}^{T} y_{it}u_{it} - p_{it}] &\leq E[\sum_{t=1}^{T-1} 6\Delta_t] + E[y_{iT}\mu_{iT}] + E[\sum_{t\in D_T\cap C_T} \mu_{it} - \gamma_t] \\ &\leq E[\sum_{t=1}^{T-1} 6\Delta_t] + E[y_{iT}\mu_{iT}] + E[\sum_{t\in C_T} \mu_{it} - \gamma_t] - E[\sum_{t\in C_T\setminus D_T} \mu_{it} - \gamma_t] \end{aligned} \tag{7}$$

For $t \in C_T$, since $\widehat{\mu}_{it}(t) \geq \widehat{\gamma}_t(t)$, we have:

$$E[\gamma_t - \mu_{it}] \quad \leq \quad E[2\Delta_t]$$

Substituting into (7):

$$\begin{aligned}
E[\sum_{t=1}^{T} y_{it}u_{it} - p_{it}] &\leq 8E[\sum_{t=1}^{T-1}\Delta_t] + E[y_{iT}\mu_{iT}] + E[\sum_{t\in C_T}\mu_{it} - \gamma_t] \\
&= 8E[\sum_{t=1}^{T-1}\Delta_t] + E[y_{iT}\mu_{iT}] + E[\sum_{t\in C_T}\mu_{it} - \widehat{\gamma}_t(T)] + E[\sum_{t\in C_T}\widehat{\gamma}_t(T) - \gamma_t] \\
&\leq 10E[\sum_{t=1}^{T-1}\Delta_t] + E[y_{iT}\mu_{iT}] + E[\sum_{t\in C_T}\mu_{it} - \widehat{\gamma}_t(T)] \qquad (8) \\
&\leq O(E[\sum_{t=1}^{T-1}\Delta_t] + E[y_{iT}\mu_{iT}]) + E[\sum_{t\in C_T}\mu_{it} - \min\{\widehat{\gamma}_t(T), \widehat{\mu}_{it}(t)\}] \\
&\leq o(E[\sum_{t=1}^{T}\eta(t)\mu_{it}]) + E[\sum_{t\in C_T}\mu_{it} - \min\{\widehat{\gamma}_t(T), \widehat{\mu}_{it}(t)\}]
\end{aligned}$$

Inequality (8) is derived by (1), and the last inequality follows by $(C1)$. The expected utility of the truthful strategy and $\mathcal{S}$ during the explore phase is equal. Therefore, by Lemma 2, the mechanism is asymptotically incentive compatible. □

Let $R(T)$ be the expected revenue of mechanism $\mathcal{M}$ between time 1 and T when all agents are truthful. Recall that the mechanism is asymptotically ex-ante efficient if

$$E[\sum_{t=1}^{T}\gamma_t] - R(T) = o(R(T)).$$

Theorem 3 *If for the learning algorithm:*

$$(C2) \qquad E[\gamma_T + \sum_{t=1}^{T-1}(\Delta_t + \eta(t)\gamma_t)] = o(E[\sum_{t=1}^{T}\gamma_t])$$

then, $\mathcal{M}$ is asymptotically ex-ante efficient.

Proof : There are three reasons why $\mathcal{M}$ may fail to be ex-ante efficient. First, during the explore phase when the item is allocated at random. The loss in potential revenue is approximately $E[\sum_{t=1}^{T}\eta(t)\mu_{it}]$. Second the error in estimation can lower payments and this loss is bounded by $O(E[\sum_{t=1}^{T-1}\Delta_t])$. Third, the payment for an item received at time T is received after time T.

Now,

$$R(T) = E[\sum_{t=1}^{T}\sum_{i=1}^{n} p_{it}] = E[\sum_{t=1}^{T-1}\sum_{i=1}^{n} y_{it}\min\{\widehat{\gamma}_t(T), \widehat{\mu}_{it}(t)\}].$$

When all the agents are truthful, for i, $y_{it} = 1$:

$$\gamma_t - \min\{\widehat{\gamma}_t(T), \widehat{\mu}_{it}(t)\} \leq \gamma_t - \min\{\widehat{\gamma}_t(T), \widehat{\gamma}_t(t)\} \leq \gamma_t - \widehat{\gamma}_t(t) \leq \Delta.$$

Therefore:

$$\begin{aligned} R(T) &\geq E[\sum_{t=1}^{T-1}\sum_{i=1}^{n} y_{it}(\gamma_t - \Delta_t)] && (9)\\ &= E[\sum_{t=1}^{T-1}(1-\eta(t))(\gamma_t - \Delta_t)] && (10)\\ &\geq E[\sum_{t=1}^{T}\gamma_t] - E[\gamma_T + \sum_{t=1}^{T-1}\Delta(t) + \eta(t)\gamma_t] \\ &= (1-o(1))E[\sum_{t=1}^{T}\gamma_t] \end{aligned}$$

Inequality (9) captures the estimation error. Inequality (10) is derived by the revenue loss during the explore phase, note that the explore phase is independent of the evolution of the utilities. The last inequality is followed by condition $(C2)$. □

3.2 Allowing agents to bid

In mechanism $\mathcal{M}$ no agent explicitly bids for an item. Whether an agent receives an item or not depends on the history of their reported utilities and the estimates that $\mathcal{M}$ forms from them. This may be advantageous when the bidders themselves are unaware of what their utilities will be. However, when agents may posses a better estimate of their utilities we would like to make use of that. For this reason we describe how to modify $\mathcal{M}$ so as to allow agents to bid for an item.

If time t occurs during an exploit phase let $\mathcal{B}_t$ be the set of the agents who bid at this time. The mechanism bids on the behalf of all agent $i \notin \mathcal{B}_t$. Denote by b_{it} the bid of agent $i \in \mathcal{B}_t$ for the item at time t. The modification of $\mathcal{M}$ sets $b_{it} = \widehat{\mu}_{it}(t)$, for $i \notin B$. Then, the item is allocated at random to one of the agents in $\arg\max_i b_{it}$.

If i is the agent who received the item at time t, let $A = \{b_{jt} | j \in \mathcal{B}_t\} \cup \{\mu_{jt} |, j \notin \mathcal{B}_t\}$. Define γ_t as the second highest value in A. Let $\widehat{\gamma}_t(T)$ to be equal to $\max_{j \neq i} b_{jk}$. The payment of agent i will be

$$p_{it} \leftarrow \sum_{k=1}^{t-1} y_{ik} \min\{\widehat{\gamma}_k(t), b_{ik}\} - \sum_{k=1}^{t-1} p_{ik}.$$

To incorporate the fact that bidders can bid for an item, we must modify the definition of truthfulness.

Definition 2 *Agent i is truthful if:*

1. *$r_{it} = u_{it}$, for all time $x_{it} = 1, t \geq 1$.*
2. *If i bids at time t, then $E[|b_{it} - \mu_{it}|] \leq E[|\widehat{\mu}_{it} - \mu_{it}|]$.*

Note that item 2 does not require that agent i bid their actual utility only that their bid be closer to the mark than the estimate. With this modification in definition, Theorems 1 and 3 continue to hold.

4 Independent and Identically-distributed Utilities

In this section, we assume that for each i, u_{it}'s are independent and identically-distributed random variables. For simplicity, we define $\mu_i = E[u_{it}], t > 0$. Without loss of generality, we also assume $0 < \mu_i \leq 1$.

In this environment the learning algorithm we use is an ε-greedy algorithm for the multi-armed bandit problem.[3] Let $n_{it} = \sum_{k=1}^{t-1} x_{it}$. For $\epsilon \in (0,1)$, we define:

$$\begin{aligned} \eta_\epsilon(t) &= \min\{1, nt^{-\epsilon}\ln^{1+\epsilon} t\} \\ \widehat{\mu}_{it}(T) &= \begin{cases} (\sum_{k=1}^{T-1} x_{ik} r_{ik})/n_{iT}, & n_{iT} > 0 \\ 0, & n_{iT} = 0 \end{cases} \end{aligned}$$

Call the mechanism based on this learning algorithm $\mathcal{M}_\epsilon(iid)$.

Lemma 4 *If all agents are truthful, then, under $\mathcal{M}_\epsilon(iid)$*

$$E[\Delta_t] = O(\frac{1}{\sqrt{t^{1-\epsilon}}}).$$

Proof : We prove the lemma by showing that for any agent i,

$$\Pr[|\mu_i - \widehat{\mu}_{it}(t)| \geq \frac{1}{\sqrt{t^{1-\epsilon}}}\mu_i] = o(\frac{1}{t^c}), \forall c > 0.$$

First, we estimate $E[n_{it}]$. There exists a constant d such that:

$$E[n_{it}] \geq \sum_{k=1}^{t-1} \frac{\eta_\epsilon(k)}{n} = \sum_{k=1}^{t-1} \min\{\frac{1}{n}, k^{-\epsilon}\ln^{1+\epsilon} k\} > \frac{1}{d} t^{1-\epsilon}\ln^{1+\epsilon} t$$

By the Chernoff-Hoeffding bound, $\Pr[n_{it} \leq \frac{E[n_{it}]}{2}] \leq e^{\frac{-t^{1-\epsilon}\ln^{1+\epsilon} t}{8d}}$.

Inequality (1) and the Chernoff-Hoeffding bound imply:

$$\begin{aligned} \Pr[|\mu_i - \widehat{\mu}_{it}(t)| \geq \frac{1}{\sqrt{t^{1-\epsilon}}}\mu_i] &= \Pr[|\mu_i - \widehat{\mu}_{it}(t)| \geq \frac{1}{\sqrt{t^{1-\epsilon}}}\mu_i \wedge n_{it} \geq \frac{E[n_{it}]}{2}] \\ &\quad + \Pr[|\mu_i - \widehat{\mu}_{it}(t)| \geq \frac{1}{\sqrt{t^{1-\epsilon}}}\mu_i \wedge n_{it} < \frac{E[n_{it}]}{2}] \\ &\leq 2e^{\frac{-\frac{1}{t^{1-\epsilon}} t^{1-\epsilon}\ln^{1+\epsilon} t\ \mu_i}{2d}} + e^{\frac{-t^{1-\epsilon}\ln^{1+\epsilon} t}{8d}} \\ &= o(\frac{1}{t^c}), \forall c > 0 \end{aligned}$$

Therefore, with probability $1 - o(\frac{1}{t})$, for all agents, $\Delta_t \leq \frac{1}{\sqrt{t^{1-\epsilon}}}$. Since the maximum value of u_{it} is 1, $E[\Delta_t] = O(\frac{1}{\sqrt{t^{1-\epsilon}}})$. □

Next, we show that $\mathcal{M}_\epsilon(iid)$ satisfies a stronger notion of individual rationality. $\mathcal{M}_\epsilon(iid)$ satisfies *ex-post individual rationality* if for any agent i, and for all $T \geq 1$:

$$\sum_{t=1}^{T} p_{it} \leq \sum_{t=1}^{T} x_{it} r_{it}.$$

[3] See [3] for a similar algorithm.

Theorem 5 *$\mathcal{M}_\epsilon(iid)$ is ex-post individually rational. Also, for $0 \le \epsilon \le \frac{1}{3}$, $\mathcal{M}_\epsilon(iid)$ is asymptotically incentive compatible and ex-ante efficient.*

Proof : We first prove ex-post individual rationality:

$$\begin{aligned}
\sum_{t=1}^{T} p_{it} &= \sum_{t=1}^{T-1} y_{it} \min\{\widehat{\gamma}_t(T), \widehat{\mu}_{it}(t)\} \\
&\le \sum_{t=1}^{T-1} y_{it} \widehat{\gamma}_t(T) \\
&\le \sum_{t=1}^{T-1} y_{it} \widehat{\mu}_{iT}(T) \\
&\le n_{it} \widehat{\mu}_{iT}(T) \\
&= \sum_{t=1}^{T-1} x_{it} r_{it}
\end{aligned}$$

The third inequality follows because the item is allocated to i at time T which implies $\widehat{\mu}_{it}(T) \ge \widehat{\gamma}_t(T)$. We complete the proof by showing that conditions $(C1)$ and $(C2)$ hold. By lemma 4, for $\epsilon \le \frac{1}{3}$:

$$E[\mu_i + \sum_{t=1}^{T-1} \Delta_t] = O(T^{\frac{1+\epsilon}{2}}) = o(T^{1-\epsilon} \ln^{1+\epsilon} T) = O(\sum_{t=1}^{T} \eta_\epsilon(t)\mu_i).$$

Therefore, $(C1)$ holds. The revenue from charging the second highest μ_{it} in each period up to time T is $T\gamma_t$. For any $\epsilon > 0$, $E[1 + \sum_{t=1}^{T-1} \Delta_t + \eta_t] = o(T)$ which implies $(C2)$. □

5 Brownian Motion

In this section, we assume for each i, $1 \le i \le n$, the evolution of μ_{it} is a reflected Brownian motion with mean zero and variance σ_i^2; the reflection barrier is 0. In addition, we assume $\mu_{i0} = 0$, and $\sigma_i^2 \le \sigma^2$, for some constant σ. The mechanism observes the values of μ_{it} at discrete times $t = 1, 2, \cdots$.

In this environment our learning algorithm estimates the reflected Brownian motion using a mean zero martingale. Define l_{it} as the last time before t that the item is allocated to i. If i has not been allocated an item yet, l_{it} is zero (recall $\mu_{i0} = 0$).

$$\eta_\epsilon(t) = \min\{1, nt^{-\epsilon} \ln^{2+2\epsilon} t\} \tag{11}$$

$$\widehat{\mu}_{it}(T) = \begin{cases} r_{il_{i,t+1}} & t < T \\ r_{il_{it}} & t = T \\ r_{il_{i,T+1}} & t > T \end{cases} \tag{12}$$

Call this mechanism $\mathcal{M}_\epsilon(\mathcal{B})$. It is not difficult to verify that the results in this section hold as long as the expected value of the error of these estimates at time t is $o(t^{\frac{1}{6}})$. However, for simplicity, we assume that the advertiser reports the exact value of μ_{it}.

We recall some well-known properties of reflected Brownian motions (see [6]).

Proposition 6 *Let $[W_t, t \geq 0]$ be a reflected Brownian motion with mean zero and variance σ^2; the reflection barrier is 0. Assume the value of W_t at time t is equal to y:*

$$E[y] = \theta(\sqrt{t\sigma^2}) \tag{13}$$

For $T > 0$, let $z = W_{t+T}$. For the probability density function of $z - y$ we have:

$$\begin{aligned} \Pr[(z-y) \in dx] &\leq \sqrt{\frac{2}{\pi T\sigma^2}} e^{\frac{-x^2}{2T\sigma^2}} && (14) \\ \Pr[|z-y| \geq x] &\leq \sqrt{\frac{8T\sigma^2}{\pi}} \frac{1}{x} e^{\frac{-x^2}{2T\sigma^2}} && (15) \\ E[|z-y| I(|z-y| \geq x)] &\leq \sqrt{\frac{8T\sigma^2}{\pi}} e^{\frac{-x^2}{2T\sigma^2}} && (16) \end{aligned}$$

Corollary 7 *The expected value of the maximum of μ_{iT}, $1 \leq i \leq n$, is $\theta(\sqrt{T})$.*

Note that in the corollary above n and σ are constant.

Lemma 8 *Suppose under $\mathcal{M}_\epsilon(\mathcal{B})$ all agents are truthful until time T, then, $E[\Delta_T] = O(T^{\frac{\epsilon}{2}})$.*

Proof : Define $X_{it} = |\mu_{i,T} - \mu_{i,T-t}|$. We first prove $\Pr[X_{it} > T^{\frac{\epsilon}{2}}] = o(\frac{1}{T^c}), \forall c > 0$. There exists a constant T_d such that for any time $T \geq T_d$, the probability that i has not been randomly allocated the item in the last $t < T_d$ step is at most:

$$\Pr[T - l_{iT} > t] < (1 - T^{-\epsilon} \ln^{2+2\epsilon} T)^t \leq e^{\frac{-t \ln^{2+2\epsilon} T}{T^\epsilon}}. \tag{17}$$

Let $t = \frac{1}{\ln^{1+\epsilon} T} T^\epsilon$. By equation (15) and (17),

$$\begin{aligned} \Pr[X_{it} > T^{\frac{\epsilon}{2}}] &= \Pr[X_{it} > T^{\frac{\epsilon}{2}} \wedge T - l_{iT} \leq t] \\ &\quad + \Pr[X_{it} > T^{\frac{\epsilon}{2}} \wedge T - l_{iT} > t] \\ &= o(\frac{1}{T^c}), \forall c > 0. \end{aligned}$$

Hence, with high probability, for all the n agents, $X_{it} \leq T^{\frac{\epsilon}{2}}$. If for some of the agents $X_{it} \geq T^{\frac{\epsilon}{2}}$, then, by Corollary 7, the expected value of the maximum of μ_{it} over these agent is $\theta(\sqrt{T})$. Therefore, $E[\max_i\{X_{it}\}] = O(T^{\frac{\epsilon}{2}})$. The lemma follows because $E[\Delta_T] \leq E[\max_i\{X_{it}\}]$. □

Theorem 9 *$\mathcal{M}_\epsilon(\mathcal{B})$ is ex- post individually rational. Also, for $0 \leq \epsilon \leq \frac{1}{3}$, $\mathcal{M}_\epsilon(\mathcal{B})$ is asymptotically incentive compatible and ex-ante efficient.*

Proof : To prove ex-post individual rationality observe that

$$\sum_{t=1}^{T} p_{it} = \sum_{t=1}^{T-1} y_{it} \min\{\widehat{\gamma}_t(T), \widehat{\mu}_{it}(t)\} \leq \sum_{t=1}^{T-1} y_{it} \widehat{\mu}_{it}(t) = \sum_{t=1}^{T-1} y_{it} r_{il_{it}} \leq \sum_{t=1}^{T} x_{it} r_{it.}$$

We complete the proof by showing the conditions $(C1)$ and $(C2)$ hold. By (13), the expected utility of each agent at time t from random exploration is $\theta(\sqrt{t\sigma^2}t^{-\epsilon}\ln^{1+\epsilon}t) = \theta(t^{\frac{1}{2}-\epsilon}\ln^{1+\epsilon}t)$. Therefore, the expected utility up to time T from random exploration is $\theta(T^{\frac{3}{2}-\epsilon}\ln^{1+\epsilon}T)$. By Lemma (8):

$$E[\mu_{iT} + \sum_{t=1}^{T-1}\Delta_t] = O(T^{1+\frac{\epsilon}{2}}).$$

For $\epsilon \leq \frac{1}{3}$, $\frac{3}{2} - \epsilon \geq 1 + \frac{\epsilon}{2}$ this yields Condition$(C1)$.

By Corollary 7, the expected value of γ_T is $\theta(\sqrt{T})$. Therefore, for any $\epsilon > 0$,

$$E[\sum_{t=1}^{T}\gamma_t] = \theta(T^{\frac{3}{2}}) = \omega(T^{\frac{3}{2}-\epsilon}\ln^{1+\epsilon}t + T^{1+\frac{\epsilon}{2}})$$

By condition $(C2)$, $\mathcal{M}_\epsilon(\mathcal{B})$ is asymptoticly ex-ante efficient. □

To apply this model to sponsored search we treat each item as a bundle of search queries. Each time step is defined by the arrival of m queries. The mechanism allocates all m queries to an advertiser and after that, the advertiser reports the average utility for these queries. The payment p_{it} is now the price per item, i.e. the advertiser pays mp_{it} for the bundle of queries. The value of m is chosen such that μ_{it} can be estimated with high accuracy.

6 Discussion and Open Problems

In this section we discuss some extensions of the mechanisms.

Multiple Slots To modify $\mathcal{M}$ so that it can accommodate multiple slots we borrow from Gonen and Pavlov [10], who assume there exist a set of conditional distributions which determine the conditional probability that the ad in slot j_1 is clicked conditional on the ad in slot j_2 being clicked. During the exploit phase, $\mathcal{M}$ allocates the slots to the advertisers with the highest expected utility, and the prices are determined according to Holmstrom's lemma ([17], see also [1]) The estimates of the utilities are updated based on the reports, using the conditional distribution.

Delayed Reports In some applications, the value of receiving the item is realized at some later date. For example, a user clicks on an ad and visits the website of the advertiser. A couple of days later, she returns to the website and completes a transaction. It is not difficult to adjust the mechanism to accommodate this setting by allowing the advertiser to report with a delay or change her report later.

Creating Multiple Identities When a new advertiser joins the system, in order to learn her utility value our mechanism gives it a few items for free in the explore phase. Therefore our mechanism is vulnerable to advertisers who can create several identities and join the system.

It is not clear whether creating a new identity is cheap in our context because the traffic generated by advertising should eventually be routed to a legitimate business. Still, one way to avoid this problem is to charge users without a reliable history using CPC.

Acknowledgment. We would like to thank Arash Asadpour, Peter Glynn, Ashish Goel, Ramesh Johari, and Thomas Weber for fruitful discussions.

References

[1] G. Aggarwal, A. Goel, and R. Motwani. Truthful auctions for pricing search keywords. *Proceedings of ACM conference on Electronic Commerce*, 2006.

[2] S. Athey, and I. Segal. An Efficient Dynamic Mechanism. *manuscript*, 2007.

[3] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning archive*, Volume 47 , Issue 2-3, 235-256, 2002.

[4] A. Bapna, and T. Weber. Efficient Dynamic Allocation with Uncertain Valuations. *Working Paper*, 2006.

[5] D. Bergemann, and J. Välimäki. Efficient Dynamic Auctions. *Proceedings of Third Workshop on Sponsored Search Auctions*, 2007.

[6] A. Borodin, and P. Salminen. Handbook of Brownian Motion: Facts and Formulae. *Springer*, 2002.

[7] R. Cavallo, D. Parkes, and S. Singh, Efficient Online Mechanism for Persistent, Periodically Inaccessible Self-Interested Agents. Working Paper, 2007.

[8] K. Crawford. Google CFO: Fraud A Big Threat. *CNN/Money*, December 2, 2004.

[9] J. Gittins. Multi-Armed Bandit Allocation Indices. *Wiley*, New York, NY, 1989.

[10] R. Gonen, and E. Pavlov. An Incentive-Compatible Multi-Armed Bandit Mechanism. *Proceedings of Third Workshop on Sponsored Search Auctions*, 2007.

[11] J. Goodman. Pay-Per-Percentage of Impressions: An Advertising Method that is Highly Robust to Fraud. *Workshop on Sponsored Search Auctions*, 2005.

[12] B. Grow, B. Elgin, and M. Herbst. Click Fraud: The dark side of online advertising. *BusinessWeek*. Cover Story, October 2, 2006.

[13] N. Immorlica, K. Jain, M. Mahdian, and K. Talwar. Click Fraud Resistant Methods for Learning Click-Through Rates. *Proceedings of First Workshop on Internet and Network Economics*, 2005.

[14] B. Kitts, P. Laxminarayan, B. LeBlanc, and R. Meech. A Formal Analysis of Search Auctions Including Predictions on Click Fraud and Bidding Tactics. *Workshop on Sponsored Search Auctions*, 2005.

[15] S. Lahaie, and D. Parkes. Applying Learning Algorithms to Preference Elicitation. *Proceedings of the 5th ACM conference on Electronic Commerce*, 2004.

[16] M. Mahdian, and K. Tomak. Pay-per-action model for online advertising. *The First International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD'07)*, 2007.

[17] P. Milgrom, Putting Auction Theory to Work. *Cambridge University Press*, 2004.

[18] D. Mitchell. Click Fraud and Halli-bloggers. *New York Times*, July 16, 2005.

[19] D. Parkes. Online Mechanisms To appear in *Algorithmic Game Theory (Nisan et al. eds.)*, 2007.

[20] B. Stone. When Mice Attack: Internet Scammers Steal Money with "Click Fraud". *Newsweek*, January 24, 2005.

[21] R. Wilson. Game-Theoretic Approaches to Trading Processes. , *Economic Theory: Fifth World Congress*, ed. by T. Bewley, chap. 2, pp. 33-77, Cambridge University Press, Cambridge, 1987.