Designing an Ad Auctions Game for the Trading Agent Competition

Patrick R. Jordan and **Michael P. Wellman**
University of Michigan
Computer Science & Engineering
2260 Hayward St
Ann Arbor, MI 48109-2121 USA

Abstract

We introduce the *TAC Ad Auctions* game (TAC/AA), a new game for the Trading Agent Competition. The Ad Auctions game investigates complex strategic issues found in real sponsored search auctions that are not captured in current analytical models. We provide an overview of TAC/AA, introducing its key features and design rationale. TAC/AA will debut in summer 2009, with the final tournament commencing in conjunction with the TADA-09 workshop.

Introduction

Since 2000, the Trading Agent Competition (TAC) series of tournaments has spurred researchers to develop improved automated bidding techniques for an array of challenging market domains. The original TAC game presented a travel-shopping scenario [Wellman et al., 2001, 2007], and subsequent games have addressed problems in supply chain management [Arunachalam and Sadeh, 2005, Eriksson et al., 2006] and market design [Niu et al., 2008]. By continually introducing new games, the TAC series engages the community in an expanded set of strategic issues bearing on trading agent design and analysis. For 2009, we have developed a fourth major game in the TAC series, in the domain of Internet advertising through sponsored search.[1]

The emergence of Internet advertising, specifically ad auctions, as a significant commercial success over the past decade [Fain and Pedersen, 2006] has led to increasing interest among academic researchers, manifest in a growing literature and a popular regular workshop on the topic. Both the commercial importance and academic interest were major motivations of introducing a new TAC game in this area. Given that bidding in keyword auctions (employing essentially the same mechanism we incorporate in the game) is a widespread current activity, the prospects for real-world implementation of ideas developed in the research competition are more direct than previous TAC games.

Despite considerable academic interest, many interesting algorithmic, bidding, and mechanism-design problems remain open [Muthukrishnan, 2008]. Designing a *realistic simulator* [Feldman and Muthukrishnan, 2008] is a central component in many of these problems. Yahoo! researchers [Acharya et al., 2007] developed the Cassini simulator in this vein. The system simulates low-level query and click behavior, publisher ranking and budget enforcement, and other aspects of the sponsored search environment. Cassini allows a rich simulation of user interaction, however the authors report some of its limitations in terms of the advertiser strategy space. For instance, advertisers were not allowed to adaptively change their bids in response to new market conditions. Perhaps most importantly, the Cassini system is not publicly accessible to the research community at large.

Another early predecessor to TAC/AA was the *Pay Per Click Bidding Agent Competition*,[2] designed and organized by Brendan Kitts as part of the ACM EC-06 Sponsored Search Workshop. Participants in this competition managed a live Microsoft AdCenter campaign for a given set of keywords over a 24-hour period. Running the competition with real money and real users over actual sponsored-search interfaces provides a maximal level of realism. In our design, however, we follow the precedent of previous TAC games in developing a simulated environment, where participants interact via a specified interface with a game server running the auctions and generating simulated market events (in this case, search user behavior). This approach provides advantages of repeatability and transparency, which are particularly important for supporting the research goals of this enterprise.

Sponsored Search

Ad auctions are used by Internet publishers to allocate and price advertising channels. Internet advertising provides a substantial source of revenue for online publishers, amounting to billions of dollars annually. Sponsored search is a popular form of targeted advertising, in which query-specific advertisements are placed alongside organic search-engine results (see Figure 1). The placement (position) of an ad for a given query, along with the cost (to the advertiser) per click (CPC), is determined through an auction process. Under cost-per-click pricing, both the publisher and advertiser bear some of the risk associated with uncertain user behavior. The use of automated auctions addresses the combinatorial problem of quoting an appropriate price (CPC) for each display slot for each distinct query. Advertisers bid for

[1] `http://aa.tradingagents.org`

[2] `http://www.biddingagentcompetition.com`

the family of keywords of interest, and competition among them determines the going CPC for each of the available slots on a query-by-query basis.

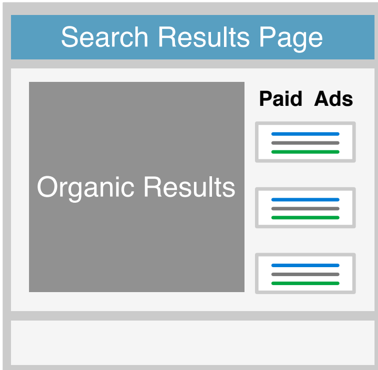

Figure 1: Typical search engine results page.

Given the salience of ad auction mechanisms, a growing number of researchers have started to investigate the *mechanism design* problem faced by search publishers, as well as the strategic problems faced by advertisers. Common to many of the early approaches are stylistic restrictions on the scenario or the full strategic space. Most of the foundational models for sponsored search analysis construct a static game of complete information for a single keyword auction [Aggarwal et al., 2006, Börgers et al., 2007, Edelman et al., 2007, Varian, 2007]. This type of analysis has provided a solid conceptual base for researchers to build upon. Significant results include equilibrium characterizations and the discovery that the auctions currently in use by publishers are not truthful. From the static models, extensions have considered dynamic variations, often evaluated through simulation [Cary et al., 2007, Lahaie and Pennock, 2007, Vorobeychik and Reeves, 2008]. TAC/AA continues in this vein by building a richer model of the environment, and follows the example of previous TAC scenarios by employing a research competition to attract experimental effort.

Designing an Ad Auction Game

The TAC/AA design attempts to include many of the interesting strategic aspects of sponsored search auctions, in a simulation framework supporting repeatability and empirical analysis. In this framework there are three types of agents as shown in Figure 2: *users*, *advertisers*, and *publishers*. The user and publisher agents are controlled by the server, while the advertiser agents are controlled by tournament participants. We discuss the behavior of each in turn below, as well as the underlying market that drives the behavior of the user and advertiser agents.

Some important elements of managing an ad campaign are not considered, such as exploration of a large keyword space for high profitability keywords, or optimizing landing page content to improve the advertiser's quality score. These issues are sacrificed not for lack of interest or value, but rather because we lack useful models to represent them.

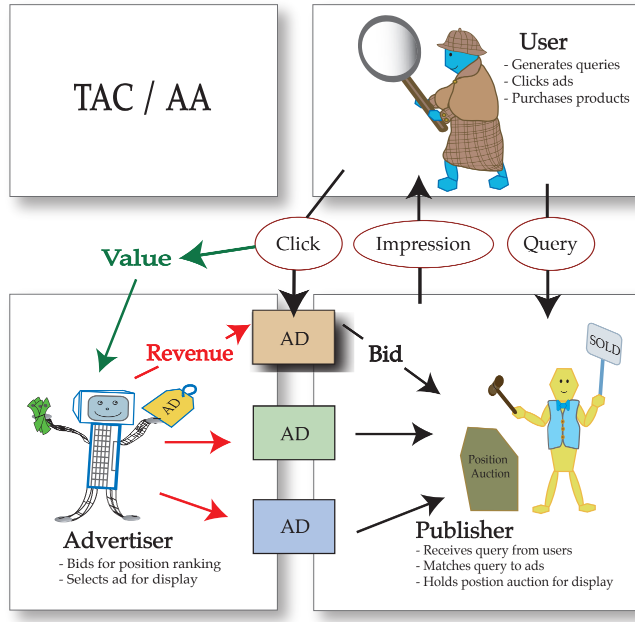

Figure 2: TAC/AA agent interaction overview.

In the process of developing TAC/AA, we identified three interesting modeling problems, not currently resolved in the sponsored search literature, central to the design of our simulation environment:

- *What drives query generation?*
- *How do advertisers derive value?*
- *Why might keyword auctions be interdependent?*

The process that generates queries is a fundamental component of ad auctions (see **Query behavior**). In addition to defining the query space, we include *query bursts*, large shifts in the number of queries, which advertisers observe in real markets. Uncertainty in the volume of queries is an important consideration for both the publishers and advertisers.

TAC/AA introduces an underlying retail market scenario (see **Defining the Market**). This market scenario defines the value of retail purchases to the advertisers. Unlike many of the earlier models, the advertiser value-per-click is not constant (see **Conversion Behavior**). This formulation imposes a keyword-value interdependency based on the query and conversion processes of user behavior. In most other models, interdependency is achieved by exogenous budgets. In reality, short-term budgets are typically not (or should not be) hard constraints. Query bursts or, in general, uncertainty in search volume are often used as justification for advertisers imposing spend limits on their campaigns. Thus, in this way spend limits can be viewed as a protectionary device that advertisers use to reduce their exposure to a large influx of low-value clicks.

A full description of the TAC/AA scenario is provided in the specification document [Jordan et al., 2009]. Here, we discuss some of the key modeling choices used in TAC/AA, providing design rationales and comparing to related literature where applicable.

Defining the Market

In the TAC/AA scenario, users search for and potentially purchase components of a home entertainment system. There is a set $\mathcal{M}$ of manufacturers in this market, each of whom produce a set of component types $\mathcal{C}$. The manufacturers and components found in TAC/AA are shown in Table 1. The set of products $\mathcal{P}$ is simply $\mathcal{M} \times \mathcal{C}$, therefore there are nine distinct products $p = (m, c)$ that are uniquely identified by their manufacturer m and component c. Advertisers represent retailers who deal in these products. Each user has an underlying preference for one of the nine products. The advertisers use the ad auctions to attract user attention to their offerings, in an attempt to generate sales.

$\mathcal{M}$	$\mathcal{C}$
Flat	TV
Lioneer	Audio
PG	DVD

Table 1: Manufacturers and components in the retail market.

Advertisers each have a distribution process that constrains their ability to deliver products to purchasing users in a timely manner. A user's decision to purchase is influenced by how constrained the advertiser is in making its deliveries. This induces a non-linearity in the value of a click (see **Conversion Behavior**) to the advertiser.

User Search Behavior

Aggarwal et al. [2008] suggest a framework for analyzing sponsored search auctions in which the *search user* takes a central role. The authors describe the need for a rich probabilistic model of user behavior, specifically once the ads are presented to the user. This corresponds to the *click* and *conversion behavior* presented in the sections that follow. We go even further and suggest that the entirety of user behavior should form the basis for analysis. This includes the definition of the query space over which that users generate queries and the frequency at which they do.

Query behavior

Search queries trigger the ad auctions that are built around them and, thus, understanding and modeling the user query process is of fundamental importance. Much of the early research in ad auctions looked at an instance of a single ad auction or a sequence of auctions all associated with a single query class. This abstracts away the interrelation among queries and the implications this has for bidding.

For instance, advertisers often use *keywords* that match multiple queries. Advertisers must reason about their values for each type of query that a single keyword matches. Moreover, the relative frequencies of queries changes dynamically over time, thus the value of the keyword changes with the distribution. This implies that query dynamics is an important consideration as well, when designing an ad auction simulation.

TAC/AA uses a state-based user model to generate this dynamic behavior (see Figure 3). Users progress through various states in order to satisfy their underlying product preferences. The user's state determines the type of query the user generates. At any given time, the population of users is divided into three broad classes: *non-searching* (NS), *searching*, and *transacted* (T). This is similar in nature to the query classifications of Jansen et al. [2008]. Non-searching users are currently inactive, generating no queries. The searching users are further divided into *informational* (IS) and *shopping* searchers. The informational searchers seek to gather information about their desired product but not to purchase. The shoppers navigate available ads and possibly transact. Shopping users are further divided by levels of search sophistication[3] (focus): low focus (level 0), intermediate (level 1), and high focus (level 2). The transacted users have satisfied their preferences and thus do not search.

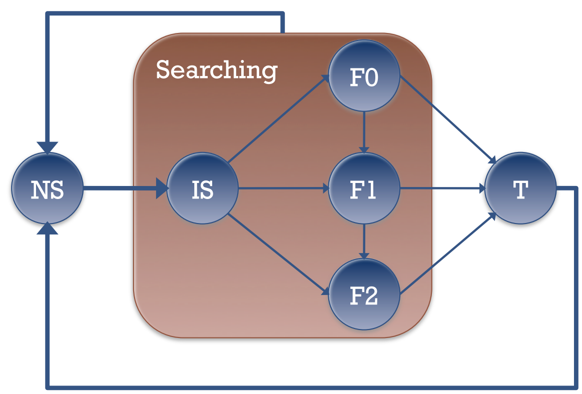

Figure 3: User state transition model. Each state also has an implicit self-loop (not shown).

A query consists of a collection of words. In our model, we consider only the six words corresponding to manufacturers and components in the home entertainment market. Each query contains at most two of these words: the user's desired manufacturer and component. For instance, a user with preference *(Lioneer, TV)* may generate a query mentioning: *Lioneer*; *TV*; both *Lioneer* and *TV*; or neither. Mentioning neither a component nor manufacturer is denoted an F0 level query. Mentioning one or the other, but not both, is denoted an F1 level query. Mentioning both component and manufacturer is denoted an F2 level query. In total, there are 16 distinct queries: 1 F0 query, 6 F1 queries and 9 F2 queries. A user with a given product preference will generate one of four queries: two possible F1 queries, and one possibility each at F0 and F2.

Each user in a searching state generates a single query per

[3]We can also think of these levels as reflecting their degree of knowledge about their own preference.

day. An F0, F1, or F2 user submits a query pertaining to its level of focus. An informational user selects among the three query types uniformly at random. If an F1 query is selected, the informational user selects between the manufacturer and component with equal probability.

Each user sub-population is modeled as a Markov chain. Most transition probabilities are stationary, with the following exceptions. To model bursts of search behavior, we provide stochastic spikes in the $NS \rightarrow IS$ transition. The transition probabilities from focused search states to state T are also non-stationary, governed by the click and conversion behavior of the user.

Click behavior

Many models have been proposed to model click behavior in users. The functional forms of the models vary, but in essence each model returns the probability that an ad at a given position will be clicked. Examples of initial models include the Edelman et al. [2007] model that assumes each position has an ad-independent click-through effect, in contrast with the Börgers et al. [2007] model that allows for an independent probability for each advertiser-position pair.

Even the Börgers et al. model is not completely general. For instance, it may be that the click probability is *dependent* on the other advertisers and the position of the other advertisers that are allocated slots. Most existing research (implicitly) adopts one of the following models for click probability:

- **Separability:** For each query, the click probability is the product of a *position* and *advertiser effect* [Aggarwal et al., 2006, Edelman et al., 2007, Börgers et al., 2007, Varian, 2007];
- **Cascade (Markovian):** For each query, each ad has a *click probability* given that the ad is viewed, as well as a *continuation probability* that the user will view the subsequent slot [Aggarwal et al., 2008, Kempe and Mahdian, 2008].

The decomposition given by the *separability model* yields a convenient form for the optimization problem the publisher solves (see **Ranking ads**), however this model does not appear to be the best predictor of click probabilities. For organic (non-sponsored) links, Craswell et al. [2008] find the *cascade model* to be the best predictor of click probabilities and argue for applicability of their results to sponsored links. The dependency of click probability on the other advertisers is termed an *externality effect*. Gunawardana and Meek [2008] analyzed these effects and found a significant *contextual effect* when *ad aggregators* were present. In general, Gunawardana and Meek's results suggest that significant externality effects exist and that the assumptions of the separability model do not hold in practice.

In contrast to the cascade model, Das et al. [2008] propose an extension of the separability model in which the user will convert from at most one of the advertisers. Like the cascade model, this also introduces a dependence on the advertisers in the higher slots.

The click model we employ in TAC/AA is a hybrid of the cascade model and the model proposed by Das et al., and also incorporates the underlying product preferences of individual search users. Users in our model proceed as in the cascade model, but stop clicking on subsequent ads when a purchase is made.

In practice, one important focus of *search engine marketing* (SEM) is selecting the ad copy or the text that is displayed in the ad. This process usually involves creating a series of ads and then testing the click-through rates of those ads, known as *split testing*. The TAC/AA click model does not incorporate text directly, however it does include a rudimentary form of ad selection. Ads take one of two forms: *targeted* and *generic*. Targeted ads emphasize a specific product, whereas generic ads do not. Ghose and Yang [2008] discuss the effects of brand and product keywords on click probability. The TAC/AA model incorporates similar effects, but in terms of ad targeting. Compared to the generic ad, users with preference matching the target of a targeted ad click with higher probability, and non-matching users are less likely to click.

Specifically, the click behavior of searching users is modeled by the following parameters:

- an advertiser effect e_q^a for each combination of advertiser a and query class q,
- a targeting effect TE which modifies the probability of clicking targeted ads depending on whether the user's preferences match the ad target,
- a promotion bonus modifying the click probability for promoted slots, and
- a continuation probability γ_q for query class q.

Given a search engine results page for query q, the user proceeds to sequentially view ads, starting from the first position. For a generic ad viewed from advertiser a, the baseline probability that the user clicks is given by e_q^a. This probability can be modified by two factors. First, the *targeting factor*, f_{target}, applies the targeting effect positively or negatively depending on whether the targeted ad selection matches user preference:

$$f_{\text{target}} = \begin{cases} 1 + TE & \text{if targeted ad, matches} \\ 1 & \text{if generic ad} \\ 1/(1+TE) & \text{if targeted ad, does not match.} \end{cases}$$

Second, the *promotion factor* f_{pro} applies a *promotion slot bonus PSB* if the ad position is a promoted slot. Promoted slots are placed in a premium location on the page (see **Slot positions**), and therefore enjoy an enhanced click rate. For a regular slot, $f_{\text{pro}} = 1$, and for a promoted slot, $f_{\text{pro}} = 1 + PSB$.

The overall click probability starts with the baseline and gets adjusted based on these factors.

$$\Pr(\text{click}) = \eta(e_q^a, f_{\text{target}} f_{\text{pro}}),$$

where

$$\eta(p, x) = \frac{px}{px + (1-p)}. \tag{1}$$

If the ad is not clicked, or clicked but no purchase is made, then the user will proceed to the next ad with continuation probability γ_q.

Conversion behavior

The purchase or conversion behavior of users can arise from various processes [Chen and He, 2006, Athey and Ellison, 2007, Cary et al., 2008, Kominers, 2008]. For example, there may be some cost associated with search for the users and the advertisers may have differentiated products and prices. In any case, these models induce some probability that the user will convert.

In TAC/AA, we describe this conversion probability in terms of inventories and backorder delays. This story is meant merely to be suggestive, just one causal explanation for the ultimate effect, which is to impose a diminishing marginal value on clicks. Our conversion model is composed of three factors. One factor is attributed to the state or type of the user. The other two factors are associated with the state of the advertiser and its product specialty, respectively.

Once an ad has been clicked-through, the shopping users will convert at different rates according to their focus levels. The probability is a function of several parameters. The baseline conversion probability is given by π_l, for $l \in \{\mathrm{F0}, \mathrm{F1}, \mathrm{F2}\}$. Higher focus level queries convert at higher rates: $\pi_{\mathrm{F2}} > \pi_{\mathrm{F1}} > \pi_{\mathrm{F0}}$.

The second factor captures an effect of constrained distribution capacity. The story is that if the advertisers sell too much product in a short period, their inventories run short and they have to put items on backorder. As a result, shoppers will be less inclined to purchase, and conversions suffer. All product sales contribute to the distribution constraint, thus rendering the queries interdependent. Let c_d be the total number of conversions over all products on day d, and W the aggregation window for distribution capacity. The distribution constraint effect is given by

$$I_d = \lambda^{\left(\sum_{i=d-1}^{d-W} c_i - C^{cap}\right)^+},$$

where C^{cap} is the critical distribution capacity, beyond which conversion rates decrease. In our scenario, advertisers are assigned one of three discrete capacity levels: $cap \in \{\mathrm{HIGH}, \mathrm{MED}, \mathrm{LOW}\}$.

Finally, we consider the effect of component specialization. For users with preference for a component matching the advertiser's specialization, the odds of converting are increased by a component specialization bonus (CSB), using the formula for odds adjustment (1). If the user matches component specialty, $f_{\text{specialization}} = 1 + CSB$, otherwise $f_{\text{specialization}} = 1$. In sum, the overall expression for conversion probability becomes

$$\Pr(\text{conversion}) = \eta(\pi_l I_d, f_{\text{specialization}}).$$

Publisher Behavior

Publishers provide the mechanism through which advertisers interact in sponsored search auctions. This includes defining the slots over which that advertisers bid, the mechanism that ranks and prices the displayed ads given the bids, and reserve prices that constrain the bids of displayed ads. The value of the slots to the advertisers and the publisher is largely determined by user behavior. This in turn requires advertisers and publishers to construct a model of user behavior in order to optimize their respective objectives. Each of these components of publisher behavior and the associated existing research are discussed subsequently.

Slot positions

When a user queries a publisher, the publisher returns a set of ads. In typical sponsored search auctions, the ads are returned in some significant order (see Figure 4). The position of the ad connotes some relative value. This relative value is inferred from the disparity in click-through rates across positions. For example, ads positioned towards the top of the results page usually have higher click-through rates, all else considered. Some search engines divide slots into two regions. One region is considered premium and is somehow set apart from the other ad slots. In TAC/AA, we distinguish two types of slots: *regular* and *promoted*. Ads in promoted slots receive an odds bonus in click-though rate.

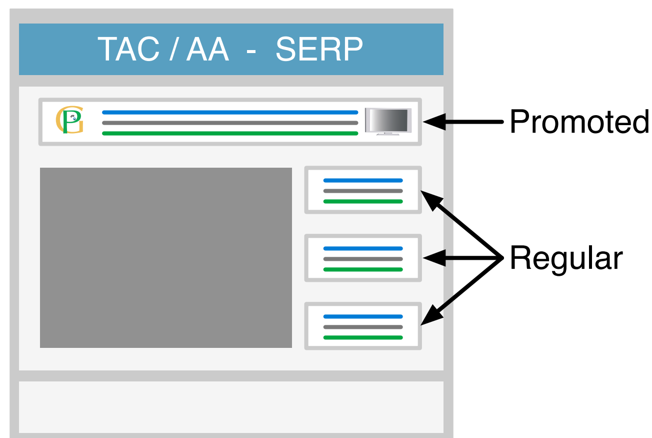

Figure 4: TAC/AA search engine results page with promoted slots.

Ranking ads

In general, each ad auction matches ads with available slots. This type of generality is appealing and general matching mechanisms have been applied to sponsored search by Aggarwal et al. [2009]. However, given the natural ordering of slots (see **Slot positions**), mechanisms that fundamentally incorporate this order are often used in practice as well as research.

Two ranking mechanisms have been dominant in the analysis of sponsored search:

- **Rank by Bid:** advertisers are ordered according to their bid b_q for a given query.
- **Rank by Revenue:** advertisers are ordered according to the product of their click-through rate and bid, $e_q b_q$, for a given query.

Lahaie and Pennock [2007] introduce a family of ranking algorithms that can interpolate between *rank-by-bid* and *rank-by-revenue*. The family is parameterized by a *squashing parameter* χ. Advertisers are ranked according to $(e_q)^\chi b_q$, which we term an advertiser's *score*. Notice that a setting of $\chi = 0$ is equivalent to rank-by-bid and a setting of $\chi = 1$ is equivalent to rank-by-revenue. The ranking method in TAC/AA uses the Lahaie and Pennock parameterization. The squashing parameter is announced at the beginning of the simulation, so that advertisers can condition their strategy on it.

Pricing clicks

In sponsored search, a slot is assigned a *cost per click* (CPC) that is determined by an auction. When a user clicks on the ad in the slot, the advertiser is charged the CPC amount. Edelman et al. [2007] describe the two basic pricing mechanisms used in sponsored search auctions.

- **Generalized first-price (GFP):** the CPC for a slot is set to the price bid by the winner of that slot.
- **Generalized second-price (GSP):** the CPC for a slot is set to the minimum price the winner of that slot needed to pay to keep the slot.

Let $b^{(i)}$ be the bid of the winner of the i^{th} position and $e^{(i)}$ be the click-through-rate of advertiser i. Using the Lahaie and Pennock parameterization, the bidder pays

$$b^{(i+1)} \left(\frac{e^{(i+1)}}{e^{(i)}} \right)^{\chi}$$

under GSP.

The auctions introduced by Overture in 1997 used GFP. Edelman et al. report one of the effects of GFP to be volatile prices. Under GFP, advertisers inevitably want to change their bid given the current setting of other-agent bids, which produces a price instability actually observed in such auctions. In practice, most publishers now use GSP, and TAC/AA adopts this pricing rule as well. With GSP, advertisers have less cause to frequently adjust prices, because they are already paying the minimum price for the slot given the other advertisers' bids.

Setting reserve prices

Reserve prices in ad auctions are used for *revenue maximization* and *ad quality control*. Abrams and Schwarz [2008] develop a framework based on the *hidden costs* advertisers impose on users. In their model, hidden costs are related to the change in future revenue due to a user clicking on an advertiser's ad. Abrams and Schwarz construct an efficient mechanism by modifying the bids by the hidden costs. Even-Dar et al. [2008] describe a set of VCG payment modifications that incorporate advertiser-specific minimum bids. One of the payment modifications offsets bids by the minimum reserve prices. Using the Abrams and Schwarz mechanism, Even-Dar et al. show that the auction is efficient and truthful. The other efficient and truthful VCG payment adjustment Even-Dar et al. introduce is *virtual values*. These virtual values essentially become *reserve scores*, where an advertiser's score is the product of its bid and click-through rate. Unlike the more general Even-Dar et al. model, the reserve price model of TAC/AA applies a uniform reserve score across advertisers for a given query. The reserve score can be converted into an advertiser-specific reserve price by adjusting for the advertiser's individual click-through rate.

Unknown user behavior

The behavior of users is not known *a priori* to publishers or advertisers. For instance, publishers may view the number of each query per day as a stochastic variable. The distribution may be influenced by many latent variables. Dealing with this type of uncertainty is an important part of a publisher's mechanism.

Recent research has explored various online algorithms for selecting allocating ad slots to advertisers given a random sequence of queries. This problem has been considered with advertiser budgets [Mehta et al., 2007, Mahdian et al., 2007, Muthukrishnan et al., 2007, Goel and Mehta, 2008] and without [Mahdian and Saberi, 2006, Abrams and Gosh, 2007].

In addition to online algorithms, publishers may try to design optimal mechanisms that use various parameterizations of user behavior. In real markets, these parameters must be learned. This learning process affects the dynamics of the auctions, which in turn affects revenue and efficiency. Wortman et al. [2007] studied the effects of this process and designed learning algorithms that maintain equilibrium during exploration.

Learning parameters is an especially important part of the publisher mechanism when the query space is large and data is sparse. In TAC/AA this is not the case, the query space is relatively small and users generate a large number of queries for each query each time period. For this reason and simplicity's sake, we just assume the publisher in TAC/AA knows advertiser-specific click probabilities, thus eliminating the need to learn click-through rates. We further assume that the ranking mechanism is fixed, so that learning more detailed user behavior is not relevant to publisher behavior. It is, of course, quite relevant to advertiser behavior.

Advertiser Strategy Space

Advertisers in sponsored search auctions face a complex problem in optimizing their ad campaigns. They contend with dynamic user behavior, uncertainty in publisher policies, and the effects of other competing advertisers. Advertisers control the content of the ads, which ads to display, the bids they place for the ads, and spend limits that bound the cost they can incur. Other aspects of campaign management are also important. For instance, optimization of the *landing page*, the page users are directed to when the click on an ad, can dramatically affect conversion rates. This has given rise to fields such as *information architecture* (IA) and *human-computer interaction* (HCI) that are devoted to improving user experience.

All of these features define the advertiser strategy space, however the TAC/AA advertisers reason over only a subset of these. Part of the motivation for excluding some features (in addition to simply limiting scope), such as *landing page optimization*, is that we expect them to be approximately strategically independent and can be studied in a decision-theoretic context apart from other strategic considerations. Features that we believe are strategically *dependent* include setting bids, choosing ads, and setting spend limits. We discuss each of these in turn over the remainder of the section.

Bidding

In TAC/AA, advertisers are given an expressive bidding language over which they are allowed to select bids. Advertisers may set a bid for any possible *query*. This contrasts with bidding languages that are actually employed by search engines where advertisers bid on *keywords*. Even-Dar et al. [2009] identify the bidding language used by TAC/AA as a *query language* and those used by the search engines as a *keyword language*.[4] In the case of a keyword language, advertisers are forced to implicitly reason about their values over a set of queries. Thus, the selection of keywords becomes a major component of the advertiser's strategy. Various natural language processing and machine learning models have been proposed that attempt to generate or select profitable keywords [Rusmevichientong and Williamson, 2006, Bartz et al., 2006, Abhishek, 2007, Chen et al., 2008]. To avoid the complexity of incorporating such concerns, we adopt a query language over the restricted domain of TAC/AA queries.

Choosing ads

In actual sponsored search auctions, advertisers generate the ads that are displayed. The content of the ad relative to the user query can have a dramatic effect on the click-through rate of the ad. Advertisers, or SEM firms managing campaigns on their behalf, typically develop ad content in an iterative manner. First, a set of candidate ads is created and submitted to the publisher for display. Then, some method of testing is used to prune ads that perform poorly. Based on the surviving ads, the advertisers generate additional candidate ads for testing and the process recurs.

The ad content in TAC/AA is specified by the inclusion, or lack thereof, of a specific product. This restricts the set of possible ads and eliminates the content creation aspect of the advertisers' strategies. However, the exploration and exploitation problem of selecting which ad to display for a given query remains.

Setting spend limits

Currently, most publishers allow advertisers to specify an *advertising budget* by which an advertiser can limit the advertising cost or spend for some period of time. Once the advertiser exceeds the limit, the constrained ads will no longer be shown.

Much of the published work on advertiser bidding strategies in dynamic, multi-keyword sponsored search auctions focuses on optimizing return while being constrained by an exogenously specified budget [Kitts and Leblanc, 2004, Zhou and Lukose, 2007, Muthukrishnan et al., 2007, Zhou et al., 2008, Zhou and Naroditskiy, 2008]. It may be the case that some advertisers do actually have a hard constraint, however we believe that in most situations the "budgets" submitted by advertisers to publishers are actually soft constraints on spending. These *daily spend limits* can be used by the advertisers to protect against a large influx of unprofitable clicks or to guard against the advertisers' uncertainty about the value of those clicks.

TAC/AA allows advertisers to specify two types of spend limits. The first type is an *aggregate* limit that binds the daily amount that an advertiser may be charged. This constrains the ads for each query class in an advertiser's campaign. The second type is a *query-class* limit, in which the daily amount charged for a specific query class is bound. Once an advertiser's daily spend amount exceeds a daily spend limit, the ad is no longer considered for inclusion in an auction.

Simulating an Advertising Campaign

The TAC/AA game simulates the daily campaigns of a set of advertisers over a horizon of two simulated months. A high level depiction of the game interaction is show in Figure 5. The game flow can be described by considering the game initialization phase and the daily tasks performed by the agents after initialization.

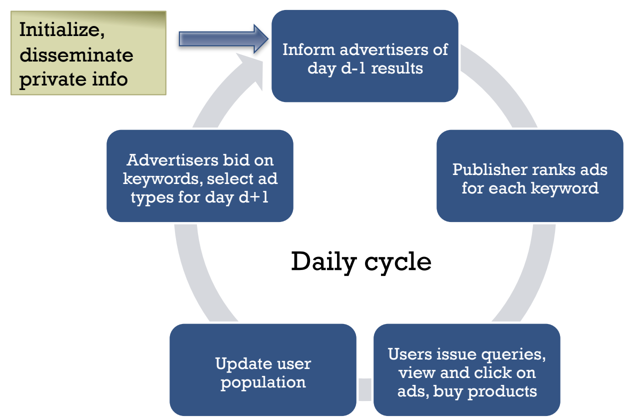

Figure 5: Cycle of activities for day d of a TAC/AA game instance.

At the beginning of a game instance, the instance-varying user, advertiser, and publisher parameter settings are drawn from their associated distributions. All users are initialized to the non-searching state, and the server simulates virtual days of user activity without advertising, to spread the population across various states. The virtual day initialization is an attempt to reduce the impact of any *cold start* anomalies. Advertisers learn their product and manufacturer specialization as well as their distribution capacity parameter (they are not told the specialties and capacities of competitors). Finally, the publisher determines and reveals the squashing parameter χ and reserve scores.

At the beginning of each day d, the daily reports summarizing day $d-1$ activity are delivered to the advertisers. The publisher executes an ad auction for each query class to determine the ad rankings and click prices. Users then issue

[4] Equivalently, one can view the TAC/AA query language as fixing a coarse partition over a large set of implicit keyword expressions.

queries, receive results, consider clicking on ads and purchasing products. The publisher monitors spend limits and reruns ad auctions as necessary. After all searching users have acted, the server updates the population based on the results of the queries, ads, and purchases. Finally, the advertisers submit their bid and ad selection updates to the publisher, for the auctions determining placement on day $d + 1$.

At the conclusion of a game, log files are produced that trace the interaction of the agents during the simulation. We provide a log file parser that allows for further post-game analysis of the traces.

TAC/AA Tournament

The TAC/AA competition will have three basic rounds: *qualifying*, *seeding*, and *finals*. During the qualifying round agents will participate in a round-robin style tournament. Agents pass the qualifying round by meeting a minimal standard for agent competence. During the seeding round agents are ranked by their average profits in a round-robin tournament. These rankings determine the bracket assignment for the finals.

The TAC/AA tournament finals will be held during the Trading Agent Design and Analysis (TADA) workshop as well as the main IJCAI conference in July 2009. The tournament will consist of multiple stages, with the particular elimination structure to be determined based on the number of entries. Following the tournament, we will release source code for the TAC/AA server, and encourage all participants to post binary versions of their agents in the TAC repository. We look forward to learning about how different teams address the strategic questions posed by TAC/AA.

Acknowledgments

Lee Callender and Ben Cassell contributed to the design and development of the TAC/AA game. Constructive suggestions and other useful feedback provided during presentations at Yahoo!, Google, and Microsoft in Fall 2008 significantly influenced several aspects of the game (we take responsibility for ultimate choices, of course). This work was supported in part by a grant from Microsoft Research.

References

V. Abhishek. Keyword generation for search engine advertising using semantic similarity between terms. In *WWW-07 Workshop on Sponsored Search Auctions*, 2007.

Z. Abrams and A. Gosh. Auctions with revenue guarantees for sponsored search. In *Workshop on Internet and Network Economics*, pages 143–154, 2007.

Z. Abrams and M. Schwarz. Ad auction design and user experience. In *Applied Economics Research Bulletin Special Issue on Auctions*, 2008.

S. Acharya, P. Krishnamurthy, K. Deshpande, T. Yan, and C.-C. Chang. A simulation framework for evaluating designs for sponsored search markets. *WWW-07 Workshop on Sponsored Search Auctions*, 2007.

G. Aggarwal, A. Goel, and R. Motwani. Truthful auctions for pricing search keywords. In *Seventh ACM Conference on Electronic Commerce*, pages 1–7, Ann Arbor, 2006.

G. Aggarwal, J. Feldman, S. Muthukrishnan, and M. Pal. Sponsored search auctions with Markovian users. In *ACM EC-08 Workshop on Ad Auctions*, Chicago, 2008.

G. Aggarwal, S. Muthukrishnan, D. Pal, and M. Pal. General auction mechanism for search advertising. In *18th International World Wide Web Conference*, Madrid, 2009.

R. Arunachalam and N. M. Sadeh. The supply chain trading agent competition. *Electronic Commerce Research and Applications*, 4:63–81, 2005.

S. Athey and G. Ellison. Position auctions with consumer search. Levine's Bibliography 122247000000001633, UCLA Department of Economics, October 2007.

K. Bartz, V. Murthi, and S. Sebastian. Logistic regression and collaborative filtering for sponsored search term recommendation. In *ACM EC-06 Workshop on Sponsored Search Auctions*, Ann Arbor, 2006.

T. Börgers, I. J. Cox, M. Pesendorfer, and V. Petricek. Equilibrium bids in auctions of sponsored links: Theory and evidence. Working paper, September 2007.

M. Cary, A. Das, B. Edelman, I. Goitis, K. Heimerl, A. R. Karlin, C. Mathieu, and M. Schwartz. Greedy bidding strategies for keyword auctions. In *Eighth ACM Conference on Electronic Commerce*, pages 262–271, San Diego, 2007.

M. Cary, A. Das, B. Edelman, K. H. Ioannis Giotis, A. R. Karlin, C. Mathieu, and M. Schwarz. On best-response bidding in GSP auctions. Working Paper 13788, National Bureau of Economic Research, February 2008.

Y. Chen and C. He. Paid placement: Advertising and search on the Internet. Working Paper No. 06-02, NET Institute, September 2006.

Y. Chen, G.-R. Xue, and Y. Yu. Advertising keyword suggestion based on concept hierarchy. In *First ACM International Conference on Web Search and Data Mining*, pages 251–260, Stanford, February 2008.

N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *First ACM International Conference on Web Search and Data Mining*, Stanford, 2008.

A. Das, I. Goitis, A. R. Karlin, and C. Mathieu. On the effects of competing advertisements in keyword auctions. Working paper, 2008.

B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review*, 97:242–259, 2007.

J. Eriksson, N. Finne, and S. Janson. Evolution of a supply chain management game for the trading agent competition. *AI Communications*, 19:1–12, 2006.

E. Even-Dar, J. Feldman, Y. Mansour, and S. Muthukrishnan. Position auctions with bidder-specific minimum prices. In *Internet and Network Economics*, LNCS 5385, pages 577–584. Springer, 2008.

E. Even-Dar, Y. Mansour, V. S. Mirrokni, S. Muthukrishnan, and U. Nadav. Bid optimization for broad match ad auctions. In *18th International World Wide Web Conference*, Madrid, April 2009.

D. C. Fain and J. O. Pedersen. Sponsored search: A brief history. *Bulletin of the American Society for Information Science and Technology*, 32(2):12–13, 2006.

J. Feldman and S. Muthukrishnan. Algorithmic methods for sponsored search advertising. In Z. Liu and C. H. Xia, editors, *Performance Modeling and Engineering*, pages 91–124. Springer, 2008.

A. Ghose and S. Yang. Analyzing search engine advertising: Firm behavior and cross-selling in electronic markets. In *17th International World Wide Web Conference*, pages 219–226, Beijing, April 2008.

G. Goel and A. Mehta. Online budgeted matching in random input models with applications to adwords. In *Nineteenth ACM-SIAM Symposium on Discrete Algorithms*, pages 982–991, San Francisco, 2008.

A. Gunawardana and C. Meek. Aggregators and contextual effects in search ad markets. In *WWW-08 Workshop on Targeting and Ranking for Online Advertising*, 2008.

B. J. Jansen, D. L. Booth, and A. Spink. Determining the informational, navigational, and transactional intent of Web queries. *Information Processing and Management*, 44: 1251–1266, 2008.

P. R. Jordan, B. Cassell, L. F. Callender, and M. P. Wellman. The Ad Auctions game for the 2009 Trading Agent Competition. Technical Report Version 0.9.3, University of Michigan, Feb. 2009. URL `http://aa.tradingagents.org`.

D. Kempe and M. Mahdian. A cascade model for externalities in sponsored search. In *ACM EC-08 Workshop on Ad Auctions*, Chicago, 2008.

B. Kitts and B. Leblanc. Optimal bidding on keyword auctions. *Electronic Markets*, 14(3):186–201, 2004.

S. D. Kominers. Dynamic position auctions with consumer search. Working paper, Harvard University, November 2008.

S. Lahaie and D. M. Pennock. Revenue analysis of a family of ranking rules for keyword auctions. In *Eighth ACM Conference on Electronic Commerce*, pages 50–56, San Diego, 2007.

M. Mahdian and A. Saberi. Multi-unit auctions with unknown supply. In *Seventh ACM Conference on Electronic Commerce*, pages 243–249, Ann Arbor, 2006.

M. Mahdian, H. Nazerzadeh, and A. Saberi. Allocating online advertisement space with unreliable estimates. In *Eighth ACM Conference on Electronic Commerce*, pages 288–294, San Diego, June 2007.

A. Mehta, A. Saberi, U. Vazirani, and V. Vizirani. Adwords and generalized on-line matching. *Journal of the ACM*, 54(5), October 2007.

S. Muthukrishnan. Internet ad auctions: Insights and directions. In *35th International Colloquium on Automata, Languages, and Programing (ICALP)*, LNCS 5125, pages 14–23, Reykjavik, July 2008. Springer.

S. Muthukrishnan, M. Pal, and Z. Svitkina. Stochastic models for budget optimization in search-based advertising. In *Internet and Network Economics*, Lecture Notes in Computer Science, pages 131–142. Springer, 2007.

J. Niu, K. Cai, S. Parsons, E. Gerding, and P. McBurney. Characterizing effective auction mechanisms: Insights from the 2007 TAC market design competition. In *Seventh International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 1079–1086, Estoril, Portugal, 2008.

P. Rusmevichientong and D. P. Williamson. An adaptive algorithm for selecting profitable keywords for search-based advertising services. In *Seventh ACM Conference on Electronic Commerce*, pages 260–269, Ann Arbor, 2006.

H. R. Varian. Position auctions. *International Journal of Industrial Organization*, 25:1163–1178, 2007.

Y. Vorobeychik and D. M. Reeves. Equilibrium analysis of dynamic bidding in sponsored search auctions. *International Journal of Electronic Business*, 6(2):172–193, 2008.

M. P. Wellman, P. R. Wurman, K. O'Malley, R. Bangera, S. de Lin, D. M. Reeves, and W. E. Walsh. Designing the market game for a trading agent competition. *IEEE Internet Computing*, 5(2):43–51, 2001.

M. P. Wellman, A. Greenwald, and P. Stone. *Autonomous Bidding Agents: Strategies and Lessons from the Trading Agent Competition*. MIT Press, 2007.

J. Wortman, Y. Vorobeychik, L. Li, and J. Langford. Maintaining equilibria during exploration in sponsored search auctions. In *Workshop on Internet and Network Economics*, pages 119–130, 2007.

Y. Zhou and R. Lukose. Vindictive bidding in keyword auctions. In *Ninth International Conference on Electronic Commerce*, pages 141–146, Minneapolis, August 2007.

Y. Zhou and V. Naroditskiy. An algorithm for stochastic multiple-choice knapsack problem and keywords bidding. In *17th International World Wide Web Conference*, Beijing, April 2008.

Y. Zhou, D. Chakrabarty, and R. Lukose. Budget constrained bidding in keyword auctions and online knapsack problems. In *Internet and Network Economics*, Lecture Notes in Computer Science, pages 566–576. Springer, 2008.