Optimization-Based Budget Pacing in eBay Sponsored Search

Qinyi Chen*
Massachusetts Institute of Technology
Cambridge, MA, USA
qinyic@mit.edu

Phuong Ha Nguyen
eBay Inc.
San Jose, CA, USA
hpnguyen@ebay.com

Djordje Gligorijevic
eBay Inc.
San Jose, CA, USA
dgligorijevic@ebay.com

ABSTRACT

In online platforms like eBay, sponsored search advertising has become instrumental for businesses aiming for enhanced visibility. However, in automated ad auctions, the sellers (ad campaigns) run the risk of exhausting their budgets prematurely in the absence of proper pacing strategies. In response to this, online platforms have been prompted to employ budget pacing strategies to maintain consistent spending patterns for their sellers. While numerous budget pacing strategies have been introduced, they predominantly stem from either empirical or theoretical perspectives, often functioning in isolation. This paper aims to bridge this gap by investigating the performance of a theoretically inspired optimization-based bid shading method, `AdaptivePacing`, within eBay's sponsored search environment and proposing variants of the algorithm tailored to real-world environments. Our findings highlight the benefits of applying theoretical pacing approaches in practical contexts. Specifically, the optimization-based `AdaptivePacing` method offers the platform flexible control over campaign spending patterns, accounts for business constraints, and suggests tailored strategies for distinct advertisers. Furthermore, when evaluating `AdaptivePacing` alongside established empirical methods, we demonstrate its practical effectiveness and pinpoint areas for further refinement.

CCS CONCEPTS

• **Information systems → Sponsored search advertising**; *Computational advertising*; • **Applied computing →** *Electronic commerce.*

KEYWORDS

online advertising, sponsored search, budget pacing, bid shading, stochastic optimization

ACM Reference Format:
Qinyi Chen, Phuong Ha Nguyen, and Djordje Gligorijevic. 2024. Optimization-Based Budget Pacing in eBay Sponsored Search. In *Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion), May 13–17, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3589335.3648331

*This work was done when Q. Chen was a research intern at eBay Inc.

WWW '24 Companion, May 13–17, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0172-6/24/05.
https://doi.org/10.1145/3589335.3648331

1 INTRODUCTION

Online marketplaces such as eBay have transformed how businesses and consumers interact, and with the industry's consistent growth, it will reach \$6.3 trillion globally (including over \$1.1 trillion in the US alone)[1]. As a leading marketplace with over 650 million monthly visits, eBay serves as a key platform for businesses to reach new audiences. Moreover, sponsored search advertising, an \$84 billion industry with current annual growth rate of 7.8%[2], is a critical tool for eBay sellers. By bidding on preferred slots or rankings, sellers can optimize visibility and drive sales of their products on a massive scale the marketplace provides. Impressions on search ads platforms are sold via automated ad auctions, where sellers specify their budget constraints, organize items into groups for which they specify targeting, as well as set maximum bids for each search keyword. Given this information, the platform then leverages an automated agent who places bids on behalf of advertisers for each impression opportunity.

However, the intricacies of such auction format bring about challenges. Sellers usually outline their targeting strategies and maximum bids in advance. Such a proactive approach, while streamlining the process, risks being blind to the sellers' budget constraints, potentially resulting in sellers depleting their budgets prematurely. This can halt sellers' momentum, sidelining them from subsequent ad auctions and causing them to miss out on valuable opportunities, especially if high-traffic or high-response periods take place after budget depletion. From the platform's perspective, ensuring a consistent and appropriate spending pattern for its sellers is also crucial for maintaining their long-term trust and partnership [14, 17].

Motivated by this realistic challenge, a growing number of budget pacing strategies have been introduced. These strategies, while distinct in their motivation, share a common goal of guiding advertisers in maintaining a steady spending pattern. Prior works on ads budget pacing can be classified into two streams, either from a practical point of view motivated by empirical considerations, or from more theoretical perspectives that seek to establish rigorous performance guarantees yet under stronger assumptions.

Empirical Viewpoints. Many different budget pacing solutions have been proposed by practitioners, who propose budget pacing methods from various angles, such as bipartite graph allocation methodologies [18], control theory [15, 21], simultaneously optimizing business metrics such as click-through-rate (CTR) or return on investment (ROI) [14, 17], or other empirical approaches [1, 22]. All of these solutions are implemented via one of the two budget pacing approaches: (i) *throttling*, where the ad campaigns are excluded from joining ad auctions probabilistically, and (ii) *bid shading*, where the advertiser's expenditure is controlled via reducing

[1]https://www.forbes.com/advisor/business/ecommerce-statistics/, accessed Nov 2023.
[2]https://www.iab.com/wp-content/uploads/2023/04/IAB_PwC_Internet_Advertising_Revenue_Report_2022.pdf, accessed Nov 2023.

their posed bids. While many of the aforementioned approaches are investigated empirically, there are no formal theoretical guarantees for the performance of these approaches.

Theoretical Viewpoints. Budget management has also aroused increasing amount of interests in the research community from a more theoretical perspective; see [2] for a comprehensive overview. These works approached the problem of budget pacing using tools such as game theory, mechanism design and learning theory, with many of them establishing game-theoretic equilibria of budgeted auctions [2, 7, 9, 10] or devised budget pacing strategies that satisfy either performance guarantees for aggregate/individual utilities or certain incentive properties [3, 5, 8, 12]. The majority of the works in this realm again studied/proposed methods that can be considered as either throttling or bid shading. While these works offer theoretical rigor that complements the prior empirical approaches, many of them rely on stringent assumptions that fail to hold in a noisy real-world environment, or relied on auction designs that are rarely used in practice [4, 13].

Our work seeks to bridge the two viewpoints above by putting theoretically inspired budget pacing approaches into an empirical context, via the eBay marketplace sponsored search. In particular, we focus on investigating the performance of an optimization-based bid shading approach first proposed in [5]; called `AdaptivePacing` (Algorithm 1). The algorithm is derived via the Lagrangian dual of an advertiser's utility optimization problem (see Section 2.3), and has been shown to be an approximate optimal strategy under ideal assumptions. By integrating the algorithm into the context of eBay's sponsored search environment, we not only assess its performance amidst noisy real-world dynamics, but also propose novel variants informed by both theoretical and empirical insights. Via answering a number of practical research questions (Section 4), we contribute in the following ways:

- We show that the optimization framework and the `AdaptivePacing` method, along with their variants, can be integrated into our real-world system and offer a number of advantages:
  1. The optimization framework gives advertisers flexible controls over their spending patterns, rather than confining to uniform or predetermined spending trajectories.
  2. It can encapsulate additional business constraints, balancing the platform's revenue with the advertisers' utilities.
  3. The theoretical insights behind this framework also suggest ways to differentiate treatment for different advertisers.
- We evaluate the performance of `AdaptivePacing` in comparison to other empirical methods like throttling and PID controllers. Drawing connections between our observations and related theoretical results, we identify actionable improvements for `AdaptivePacing`; see Section 4.2.

While this work focuses on the pay-per-click advertising model, these contributions also apply to any other advertising models.

2 PRELIMINARIES AND METHODOLOGY

In this section, we introduce the setup of an sponsored search environment. We then formulate the optimization problem for ad campaigns given the auction mechanism, and describe the `AdaptivePacing` algorithm (which is first introduced in [5] and adapted for our sponsored search setup). `AdaptivePacing` serves as the basis for our development and testing in subsequent sections.

2.1 Background on sponsored search

In a sponsored search program, there are a number of search activities (impressions) $i \in \mathcal{I}_t$ being initiated at any given time $t \in [T]$. Whenever a user initiates a search, each ad campaign of the sponsored search program, denoted by $k \in [K]$, is considered for retrieval to an internal ad auction that determines the allocation of the sponsored placements for the search result page.

For each impression i, campaign k determines a maximum bid it can post, denoted by $v_{k,i} \in [0, \bar{v}]$. Retrieval is then performed based on campaigns' targeting strategies which can include manual and auto targeting as well as their extensions. If campaign k's targeting strategy does not align with impression i, we simply let $v_{k,i} = 0$. Here, the maximum bid $v_{k,i}$ can be viewed as a proxy for the ad campaign's valuation of the ad slot.

In addition to its maximum bid for each impression, the ad campaign also provides the platform with (i) its *total budget*, denoted by B_k, which is the total monetary value to be spent throughout the horizon; and (ii) its *target spending curve*, captured by by $\boldsymbol{\rho}_k \in \Delta_T$, where $\rho_{k,t}$ is the percentage of budget that campaign k wishes to allocate to round t. The target spending curve determines the rate at which the campaign wishes to spend its budget over time. Some common examples include the uniform spending curve, the traffic curve, the CTR curve, etc. (See Sections 4.1.1 and 4.2.2 for experiments on different spending curves.)

Second-price multi-slot ads auction. For each search activity/impression i, the sponsored search program conducts a second-price multi-slot ads auction to determine the allocation of slots. External competition, including other ad programs or ad deals, can also enter this auction and thus impact campaigns' strategies. Each campaign in the sponsored search program should submit a bid $b_{k,i} \in [0, v_{k,i}]$, bounded by its maximum bid value. If $b_{k,i} = 0$, campaign k opts out of this auction. Having collected the bids from all campaigns, the platform computes an *ad expected value* for each campaign: $r_{k,i} = b_{k,i} p_{k,i}$, which is determined using campaign k's bid value $b_{k,i}$ and its response probability $p_{k,i}$ (i.e., click-through rate for cost-per-click goal type). The campaigns are then ranked based on their ad expected values.

If impression i offers N sponsored slots, the campaigns with the top N ad expected values would be allocated a slot in descending order. Each winning campaign is charged a clearing price $c_{k,i}$ if a user response is recorded (i.e., click). Here, the clearing price is computed based on the ad expected value of the campaign that immediately follows, defined as $d_{k,i} = \max_{i': r_{k,i'} < r_{k,i}} (r_{k,i'})$. The clearing price is subsequently determined as $c_{k,i} = d_{k,i}/p_{k,i}$, which is inherently less or equal to the bid $b_{k,i}$.[3]

[3]For simplicity, in the rest of Section 2, we consider single-slot auctions when formulating the campaign's optimization problem and pacing algorithm. Nonetheless, the same optimization framework would readily extend to a multi-slot setup, as discussed by [12]. In our real-world experiments in Section 4, we also consider multiple ad slots.

2.2 Campaign's optimization problem

At the auction for impression i, each campaign k posts a bid $b_{k,i} \in [0, v_{k,i}]$ and has response probability $p_{k,i}$. With slight abuse of notation, we let $d_{k,i}$ be the highest ad expected value among the rest of the campaigns: $d_{k,i} = \max_{k' \neq k} b_{k',i} \cdot p_{k',i}$. Campaign k also has a budget B_k which constrains its maximum amount of spending throughout the horizon.

Campaign k needs to determine $x_{k,i} \in \{0,1\}$, i.e., whether it wishes to win impression i. Note that if campaign k is not constrained by budget, it would always wish to win the impression. However, with the budget constraint in mind, campaign k needs to focus on winning impressions that yield the highest utilities. The expected utility that campaign k receives from impression i is

$$x_{k,i}\left(v_{k,i} - c_{k,i}\right) p_{k,i} = x_{k,i}\left(v_{k,i}p_{k,i} - d_{k,i}\right) \tag{1}$$

where $v_{k,i} - c_{k,i}$ captures the utility gained by campaign k if it is shown in search i and receives a user response, while $p_{k,i}$ is the user response probability. The equality in (1) follows from the definition of the clearing price ($c_{k,i} = d_{k,i}/p_{k,i}$). Similarly, we can write the expected spending of campaign k on impression i as follows:

$$x_{k,i}c_{k,i}p_{k,i} = x_{k,i}d_{k,i}. \tag{2}$$

Given Equations (1) and (2), a campaign k that seeks to maximize its total expected utility throughout the horizon subject to its budget constraint can thus solve the following optimization problem:

$$\begin{aligned} \max_{x_{k,i}\in\{0,1\}} & \sum_{t=1}^{T}\sum_{i\in\mathcal{I}_t} x_{k,i}\left(v_{k,i}p_{k,i} - d_{k,i}\right) \\ \text{s.t.} & \sum_{t=1}^{T}\sum_{i\in\mathcal{I}_t} x_{k,i}d_{k,i} \le B_k \end{aligned} \tag{3}$$

2.3 The `AdaptivePacing` algorithm

The campaign's optimization problem in (3) motivates the design of the `AdaptivePacing` algorithm (detailed in Algorithm 1), which was first introduced by [5]. This algorithm has been tailored in our study to suit the context of sponsored search, as discussed in Section 2.1. Below, we provide an overview of the fundamental concepts that underpin the design of `AdaptivePacing`.

Algorithm 1 `AdaptivePacing`

Input: Total budget B_k, target spending rate $\boldsymbol{\rho}_k$, step size $\epsilon_{k,t}$

(1) Initialize pacing multiplier $\mu_{k,t} = 0$

(2) For t in $1, \ldots, T$

(a) Whenever campaign k joins auction for impression $i \in \mathcal{I}_t$, with maximum bid $v_{k,i}$
- Post bid $b_{k,i} = \frac{v_{k,i}}{1+\mu_{k,t}}$
- Realized spending $\tilde{z}_{k,i} = b_{k,i}$ if campaign k wins the impression and it gets clicked; $\tilde{z}_{k,i} = 0$ otherwise.

(b) Compute total realized spending $\tilde{z}_{k,t} = \sum_{i\in\mathcal{I}_t} \tilde{z}_{k,i}$

(c) Update the pacing multiplier

$$\mu_{k,t+1} = \mu_{k,t} - \epsilon_{k,t}(\rho_{k,t}B_k - \tilde{z}_{k,t})$$

To solve Problem (3) in a real-time fashion, we start by formulating the Lagrangian of the optimization problem in (3) by introducing $\mu_k \ge 0$ as the dual variable associated with the budget constraint. We then write the Lagrangian dual problem of (3) as the following:

$$\begin{aligned} & \inf_{\mu_k\ge 0} \max_{x_{k,i}\in\{0,1\}} \sum_{t=1}^{T}\sum_{i\in\mathcal{I}_t} x_{k,i}\left(v_{k,i}p_{k,i} - d_{k,i}\right) + \mu_k \cdot (B_k - \sum_{t=1}^{T} x_{k,t}d_{k,t}) \\ = \; & \inf_{\mu_k\ge 0} \max_{x_{k,i}\in\{0,1\}} \sum_{t=1}^{T}\Big[\sum_{i\in\mathcal{I}_t} x_{k,i}\left(v_{k,i}p_{k,i} - (1+\mu)d_{k,i}\right) + \mu_k\rho_{k,t}B_k\Big] \end{aligned}$$

Note that to solve the inner optimization problem, the optimal solution is to let $x_{k,i} = \mathbb{1}\{v_{k,i}p_{k,i} \ge (1+\mu_k)d_{k,i}\}$. That is, the campaign wants to win all auctions t such that $v_{k,i}p_{k,i} \ge (1+\mu_k)d_{k,i}$. This is achieved by having campaign k post bid

$$b_{k,i} = \frac{v_{k,i}}{1+\mu_k}.$$

We call μ_k the *pacing multiplier* of campaign k, as it directly regulates the extent of pacing by modifying the bid value.

Having solved the inner optimization problem, we can then rewrite the Lagrangian dual problem of (3) as the following:

$$\inf_{\mu_k\ge 0} \sum_{t=1}^{T}\Big[\sum_{i\in\mathcal{I}_t}\left(v_{k,i}p_{k,i} - (1+\mu)d_{k,i}\right)^+ + \mu\rho_{k,t}B_k\Big], \tag{4}$$

where $y^+ := \max(y, 0)$. Given the Lagrangian dual problem in (4), we next solve for the pacing multiplier μ_k such that the above objective is minimized. Since we do not have prior information on observed values $v_{k,i}$ and competing bids $d_{k,i}$, the adaptive pacing algorithm approximates the optimal μ_k using a subgradient descent method, where we maintain a pacing multiplier $\mu_{k,t}$ at the end of each round, and keep updating its value in the following way.

In particular, consider

$$\phi_{k,t}(\mu_k) := \sum_{i\in\mathcal{I}_t}\left(v_{k,i}p_{k,i} - (1+\mu)d_{k,i}\right)^+ + \mu\rho_{k,t}B_k,$$

with the following subgradient:

$$\partial\phi_{k,t}(\mu) = \rho_{k,t}B_k - \sum_{i\in\mathcal{I}_t} d_{k,t}\mathbb{1}\{v_{k,t}p_{k,i} \ge (1+\mu)d_{k,t}\} = \rho_{k,t}B_k - z_{k,t}(\mu),$$

where $z_{k,t}(\mu) := \sum_{i\in\mathcal{I}_t} d_{k,t}\mathbb{1}\{v_{k,t}p_{k,i} \ge (1+\mu)d_{k,t}\}$ is campaign k's expected amount of expenditure if it posts bid $b_{k,i} = \frac{v_{k,i}}{1+\mu}$ for all auctions $i \in \mathcal{I}_t$ at the tth round. This prompts us to perform the following (sub)gradient descent scheme on the dual multiplier $\mu_{k,t}$:

$$\mu_{k,t+1} = \mu_{k,t} - \epsilon_{k,t}(\rho_{k,t}B_k - \tilde{z}_{k,t}),$$

where $\tilde{z}_{k,t}$ is the actual realized expenditure incurred by campaign k during the tth round, and $\epsilon_{k,t}$ is the update step size.

The theoretically sound `AdaptivePacing` algorithm would serve as the main building block of our study. In subsequent sections, we will showcase its practical effectiveness within a real-world system and adapt `AdaptivePacing` to accommodate a range of business cases. Additionally, we also investigate its empirical links to well-established methods in control theory and pursue further enhancements inspired by these connections.

3 EBAY SPONSORED SEARCH TEST BED

In this section, we detail the production simulation test bed for a sponsored search program in an online marketplace, which is the simulation environment we will adopt for all of our experiments in Section 4. We also introduce the key business metrics that we use for evaluation of budget pacing algorithms.

3.1 Sponsored search test bed

Sponsored search test bed is a counterfactual simulator based on historical records of users' search requests, simulating the entire product environment for a duration of a day, which is a typical budget duration of sponsored search campaigns' budgets [19]. The sponsored search gym is an environment that scales by subdividing the simulation period into small time segments within which each search request and ad auction is treated independently. Real-time budget signal estimation stands as the optimal approach, but its implementation may slow down research progress and, more importantly, could destabilize the practical system due to the heavy computational load. Consequently, adopting near-real-time updates with a one-minute resolution emerges as a judicious compromise between computational complexity and modeling precision.

The process of retrieval based on targeting input, an essential component in practical scenarios, is intricate and time-intensive. It necessitates activities like sorting by relevance and other critical metrics. To emulate this targeting-based retrieval task, a unique targeting set is constructed for each search query, sourced from logged recall sets. The bid value associated with an item within a campaign is contingent on its targeting strategy, particularly concerning keywords, which might exhibit overlaps. Consequently, the calculation of the *ad expected value* accords precedence to the ad group with the highest bid, akin to the production system's approach. Probability of user response is generated based on available features, while user response necessary to generate spend is simulated on the principle of counterfactual modeling.

The quality of the test bed was reported earlier in [19], comparing it to the original data and results based on naive simulations. Although certain disparities in the simulation results vis-à-vis real traffic may arise from the approximations inherent to the test bed, the reported results show that the principal components of the sponsored search gym, notably response probability generation, result in a notable enhancement across key performance metrics.

3.2 Business metrics

Our evaluation of budget pacing approaches involves assessing a number of key system-level performance metrics. The main metrics we consider include the number of impressions (**Imps**) won by the campaigns in our sponsored program; the number of clicks (**Clks**); total ad revenue (**Rev**); average click-through rate (**CTR**); average cost-per-click (**CPC**), a seller-oriented metric that evaluates their cost; surface rate (**SR**), which is the percentage of sponsored slots won by campaigns in the sponsored search program (recall that other ad programs/deals can also join the auctions and compete for visibility). Finally, another central metric we assess is the pacing error (**PE**), defined as: PE $= (1/T) \times \sum_{t=1}^{T} \frac{|pS_t - pT_t|}{pT_t}$, with pS_t and pT_t as the fraction of total spend and traffic at t-th round. This metric measures smoothness of system-level spend over a day using the traffic curve as a reference. A lower pacing error is more ideal.

4 PRACTICAL RESEARCH QUESTIONS AND EXPERIMENT RESULTS

In this section, we aim to reconcile theory and practice by evaluating the performance of the optimization-based `AdaptivePacing` method and its variants, and making comparisons with other commonly adopted budget pacing approaches. We leverage eBay's sponsored search test bed detailed in Section 3.1, and evaluate the performance of our algorithms using key metrics in Section 3.2. Based on these, we address a number of research questions of practical significance to online platforms.

In the following experiments, all of the results are expressed as percentage variations relative to the metrics observed under no-pacing (that is, each campaign k always posts maximum bid $v_{i,t}$ whenever it joins an auction for impression i, until its budget gets depleted). We let each round t span one minute and the entire time horizon is set to be $T = 1440$, which is the duration of one day. The default target spending curve used by `AdaptivePacing` and its variants is the traffic curve, where $\rho_{k,t} = |\mathcal{I}_t| / \sum_{t=1}^{1440} |\mathcal{I}_t|$ represents the fraction of traffic during the t-th minute. The default step size is set to be $\epsilon_{k,t} = 0.01$ unless stated otherwise.

4.1 Optimization-based budget pacing in a real-world system

We start by evaluating the effectiveness of the optimization-based `AdaptivePacing` method in our real-world test-bed.

4.1.1 ***Does the optimization-based Adaptive Pacing approach work in a real-world system?*** As alluded in our prior discussion, real-world systems do not really satisfy the many assumptions that are deemed crucial to the performance guarantees of the Adaptive Pacing Algorithm, as stated in [5]. In particular, [5] established that (i) in a stationary environment where $v_{k,i}$ and $d_{k,i}$ are both drawn independently from fixed distributions, `AdaptivePacing` would enjoy near-optimal performance in the long run; (ii) in an adversarial setting, the algorithm is shown to ensure that the campaign would receive $B_k/\bar{v}T$ of the optimal utility in hindsight as $T \to \infty$ (see Theorem 3.3 in [5]). However, in a real-world online marketplace with arbitrary arrivals, the insights one can obtain from the above theoretical results remain limited. This is because (i) the environment is in no way stationary as the campaigns provide maximum bids in an arbitrary fashion; (ii) we may assume that a majority of campaigns have small budgets or receive small number of clicks due to high competition in the environment, making the theoretical ratio under adversarial setting $B_k/\bar{v}T$ arbitrarily small; (iii) the budgets of all campaigns are typically reset daily or weekly, meaning that we are working with a non-asymptotic time horizon.

Having these in mind, we first evaluate the performance of the vanilla `AdaptivePacing` (Algorithm 1) in eBay's test bed, under different target spending curves. To better evaluate its performance, we additionally consider a throttling approach from [19] (referred to as `Throttling`) as our baseline, which probabilistically determine whether to let campaigns enter auctions and aims for uniform spending for all campaigns (see Appendix A for a description of this method and a detailed comparison between throttling and our bid shading approach). All results are recorded as percentage variations compared to metrics attained under no-pacing.

Table 1 shows that `AdaptivePacing` outperforms no-pacing and `Throttling` across several crucial metrics, such as impressions, clicks, and surface rate, indicating that campaigns are utilizing their budgets more effectively and appearing more frequently in search results. CPC also decreases significantly, further enhancing advertiser utilities. We observe a decrease in total ad revenue precisely because `AdaptivePacing` always lets campaigns post bids less than their maximum bids, which reduces the clearing price (as seen in CPC) under the second-price auction mechanism. Finally, in terms of the pacing error, `AdaptivePacing` manages to reduce pacing error by as high as 18% when the traffic curve is adopted. Conversely, while `Throttling` minimizes pacing error, it does so at the cost of other vital metrics (see Appendix A for a detailed evaluation of throttling.) We also invite readers to see Appendix B for illustrations of business metrics under `AdaptivePacing` and `Throttling` over the entire horizon, which shows how `AdaptivePacing` lets advertisers seize opportunities whenever they arise.

In Table 1, the performance of `AdaptivePacing` is evaluated under different target spending curves, captured by $\boldsymbol{\rho}_k$. We have tried (i) the traffic curve, where $\rho_{k,t}$ is the proportion of the amount of traffic during the t-th minute; (ii) the uniform spending curve, where $\rho_{k,t} = 1/1440$, which aims to spend uniformly in each minute; (iii) the CTR curve, where $\rho_{k,t}$ is set to be proportional to the CTR curve from prior data. We can observe from Table 1 that different target spending curves lead to different outcomes, and no single target spending curve would dominate from the advertisers' perspective. For example, the traffic curve leads to the most improvements in terms of the number of impressions, clicks, and pacing error, while the campaigns saves the most spending under the CTR curve. `AdaptivePacing` enables the campaigns (advertisers) to flexibly select any target spending curve best suited to their needs. See Section 4.2.2 for an extended discussion on how different target spending curves influence the campaigns' spending patterns.

Table 1: Performance of `AdaptivePacing` (w. different target spending curves) and `Throttling` [19]. Results are shown as percentage variations relative to no-pacing.

Spending Curve	Imps	Clks	Rev	CTR	CPC	SR	PE
AdaptivePacing							
Traffic	**6.21%**	**5.48%**	-5.39%	-0.50%	-10.20%	**6.01%**	-18.72%
Uniform	5.51%	5.03%	-6.26%	**-0.26%**	-10.55%	5.79%	-10.74%
CTR	5.46%	4.90%	-6.43%	-0.35%	**-10.57%**	5.89%	-9.85%
Throttling							
Uniform	-0.45%	-1.10%	**-1.89%**	-0.88%	-0.28%	-7.76%	**-51.62%**

Despite the absence of key assumptions in a noisy real-world setting, the performance of `AdaptivePacing` aligns well with the theoretical results presented in [5], and confirm that `AdaptivePacing` can substantially enhance the benefits for campaigns (advertisers) by offering them an increased number of impressions/clicks with reduced costs due to a drop in the clearing price. However, note that this also presents an inherent tradeoff between the utilities campaigns receive and the revenue garnered by the platform. The drop in campaigns' spending also implies a potential revenue loss for the platform, which might deter the platform from implementing the `AdaptivePacing` approach. A better mechanism might need to be introduced to incentivize the platform to employ budget pacing. An ideal system would grant the platform the capability to calibrate the perceived trade-offs, striking a balance between its own business objectives (revenue) and maximizing advertisers' utilities. This will be investigated in the subsequent section.

4.1.2 *How to balance the platform's and advertisers' interests?*

`AdaptivePacing` (Algorithm 1) is designed with the goal of maximizing the overall gain of individual ad campaigns, as suggested by the optimization framework in Section 2.3. However, as seen in Section 4.1.1, `AdaptivePacing` would lead to a decrease in the platform's revenue as high as over 6%, which conflicts with the platform's primary objective of maximizing its own revenue.

To reconcile the platform's objective with advertiser utilities, we present an alternative optimization problem for campaign k, which introduces an additional constraint to ensure that the platform would receive a reasonable share of revenue after budget pacing:

$$\begin{aligned} \max_{x_{k,i}\in\{0,1\}} \quad & \sum_{t=1}^{T}\sum_{i\in\mathcal{I}_t} x_{k,i}\left(v_{k,i}p_{k,i} - d_{k,i}\right) \\ \text{s.t.} \quad & \sum_{t=1}^{T}\sum_{i\in\mathcal{I}_t} x_{k,i}d_{k,i} \le B_k \\ & \sum_{t=1}^{T}\sum_{i\in\mathcal{I}_t} x_{k,i}d_{k,i} \ge \alpha_k \cdot B_k \end{aligned} \tag{5}$$

where the second constraint is called the *minimum-spending constraint*, and $\alpha_k \in [0,1]$ is the minimum percentage of budget that the platform would like campaign k to at least spend throughout the horizon. Here, we allow the platform to freely determine the value of α_k based on its business need.

Using similar derivations as in Section 2.3, we let $\mu_k, \gamma_k \ge 0$ be the dual variables associated with the budget constraint and the minimum-spending constraints. Solving the Lagrangian dual problem problem of (5) gives that campaign k should post bid

$$b_{k,i} = \frac{v_{k,i}}{1 + (\mu_k - \gamma_k)^+} \tag{6}$$

whenever it bids for impression i. Here, μ_k is the *pacing multiplier* we had previously in Section 2.3, and we call γ_k the *spending multiplier* that additionally regulates the extent of spending.

We can again update both multipliers using the subgradient descent method as follows:

$$\mu_{k,t+1} = \left(\mu_{k,t} - \epsilon_{k,t}\left(\rho_{k,t}B_k - \tilde{z}_{k,t}\right)\right)^+$$
$$\gamma_{k,t+1} = \left(\gamma_{k,t} - \epsilon'_{k,t}\left(\tilde{z}_{k,t} - \alpha_k \cdot \rho_{k,t}B_k\right)\right)^+$$

where $\epsilon_{k,t}$ is the step size that controls our rate of pacing, while $\epsilon'_{k,t}$ is the step size that controls our rate of stimulating spending. This then gives a variant of `AdaptivePacing` that additionally allows the platform to ensure its business objective (revenue) remains at a desired level, while maximizing the advertisers' utilities. We call this algorithm `AdaptivePacing-SpendingPenalty` (Algorithm 2).

Remark 4.1 (**Connection with choosing the right spending curve.**). *We remark that* `AdaptivePacing-SpendingPenalty` *(Algorithm 2) is in fact equivalent to vanilla* `AdaptivePacing` *(Algorithm 1) that adopts a carefully crafted target spending curve.*

Algorithm 2 `AdaptivePacing-SpendingPenalty`

Input: Total budget B_k, target spending rate $\boldsymbol{\rho}_k$, minimum spending percentage α_k, step size $\epsilon_{k,t}$

(1) Initialize pacing and spending multipliers $\mu_{k,t}, \gamma_{k,t} = 0$

(2) For t in $1, \ldots, T$

  (a) Whenever campaign k joins auction for impression $i \in \mathcal{I}_t$, with maximum bid $v_{k,i}$
  - Post bid $b_{k,i} = \frac{v_{k,i}}{1+(\mu_{k,t}-\gamma_{k,t})^+}$
  - Realized spending $\tilde{z}_{k,i} = b_{k,i}$ if campaign k wins the auction and gets clicked; $\tilde{z}_{k,i} = 0$ otherwise.

  (b) Compute total realized spending $\tilde{z}_{k,t} = \sum_{i \in \mathcal{I}_t} \tilde{z}_{k,i}$

  (c) Update the pacing and spending multipliers

$$\mu_{k,t+1} = \left(\mu_{k,t} - \epsilon_{k,t}\left(\rho_{k,t}B_k - \tilde{z}_{k,t}\right)\right)^+$$

$$\gamma_{k,t+1} = \left(\gamma_{k,t} - \epsilon'_{k,t}\left(\tilde{z}_{k,t} - \alpha_k \cdot \rho_{k,t}B_k\right)\right)^+$$

To see that, note that if we take $\epsilon'_{k,t} = c \cdot \epsilon_{k,t}$ for some constant $c > 0$, and let $\eta_{k,t} = \mu_{k,t} - \gamma_{k,t}$, applying `AdaptivePacing-SpendingPenalty` *is equivalent to having campaign k post bid $b_{k,i} = \frac{v_{k,i}}{1+\eta_{k,t}}$ for impression $i \in \mathcal{I}_t$ at round t, and making the following update:*

$$\begin{aligned}\eta_{k,t+1} &= \mu_{k,t+1} - \gamma_{k,t+1} \\ &= \left(\mu_{k,t} - \epsilon_{k,t}\left(\rho_{k,t}B_k - \tilde{z}_{k,t}\right)\right) - \left(\gamma_{k,t} - \epsilon'_{k,t}\left(\tilde{z}_{k,t} - \alpha_k \cdot \rho_{k,t}B_k\right)\right) \\ &= \left(\mu_{k,t} - \gamma_{k,t}\right) - \epsilon_{k,t}\left((1 + c \cdot \alpha_k) \cdot \rho_{k,t}B_k - (1+c)\tilde{z}_{k,t}\right) \\ &= \eta_{k,t} - (1+c)\epsilon_{k,t} \cdot \left(\frac{1 + c \cdot \alpha_k}{1+c} \cdot \rho_{k,t}B_k - \tilde{z}_{k,t}\right)\end{aligned}$$

which is equivalent to `AdaptivePacing` *with a different step size $\tilde{\epsilon}_{k,t} = (1+c)\epsilon_{k,t} = \epsilon_{k,t} + \epsilon'_{k,t}$ and a different target spending curve*

$$\tilde{\rho}_{k,t} = \frac{1 + c \cdot \alpha_k}{1+c} \cdot \rho_{k,t}\,.$$

Note that as $c \to 0$, `AdaptivePacing-SpendingPenalty` *exactly approaches vanilla* `AdaptivePacing`.

Nonetheless, the target spending curve $\tilde{\rho}_{k,t}$ here is in general more difficult to determine in practice. Our optimization-based formulation in (5)*, on the other hand, allows us to observe/control the rates of pacing and spending in a more interpretable fashion. One can immediately tell how much the campaign is overspending (underspending) at any time by directly observing the magnitude of the pacing (spending) multiplier. For instance, if at the t-th minute campaign k has multipliers $\mu_{k,t} > 0, \gamma_{k,t} = 0$, it is evident that this campaign is overspending at that point; and vice versa.*

In Table 2, we evaluate `AdaptivePacing-SpendingPenalty` under varied step sizes for pacing ($\epsilon_{k,t} = \epsilon$) and spending ($\epsilon'_{k,t} = \epsilon'$) respectively, which encapsulate the platform's prioritization between spending and pacing. For the purpose of comparison, for each campaign, we set the minimum spend percentage α_k to be the percentage of the budget spent by campaign k under no-pacing. In reality, the platform is free to determine the minimum spend percentage based on how much it cares about maintaining its revenue. It can be seen from Table 2 that when we take $\epsilon_{k,t} = 0.01$ and $\epsilon'_{k,t} = 0.1$ (i.e., when the platform places more priority on maintaining its own revenue than budget pacing), the total spending of campaigns got significantly increased compared to `AdaptivePacing`.

Table 2: Performance of `AdaptivePacing-SpendingPenalty` under different update step sizes for pacing multipliers $\epsilon_{k,t} = \epsilon$ and for spending multipliers $\epsilon'_{k,t} = \epsilon'$.

ϵ	ϵ'	Imps	Clks	Rev	CTR	CPC	SR	PE
0.1	0.1	**5.63%**	**5.13%**	-3.50%	-0.32%	-8.20%	**5.58%**	-20.08%
0.01	0.01	3.95%	3.20%	-5.15%	-0.64%	-8.03%	3.51%	**-37.98%**
0.01	0.1	4.06%	3.67%	**-1.16%**	**-0.29%**	-4.88%	4.30%	-14.09%
0.1	0.01	3.77%	2.75%	-8.27%	-0.89%	**-10.42%**	3.63%	-37.26%

4.1.3 *How do we differentiate treatments for different types of campaigns?*

In all of the experiments above, we have applied the same pacing methods to all campaigns simultaneously. This raises the following question: Would differentiating treatments for different campaigns improve our business metrics? For example, we have observed a drop in platform's revenue due to `AdaptivePacing` decreasing clearing prices for all campaigns. Can we restrict pacing only to some campaigns during certain time periods to prevent clearing prices from dropping too much?

Table 3: Performance of heuristics of `AdaptivePacing` that ONLY adopt budget pacing under certain conditions.

Conditions	Imps	Clks	Rev	CTR	CPC	SR	PE
`AdaptivePacing-Budget`							
$B_k \leq 5000$	5.18%	4.72%	**-3.82%**	**-0.27%**	-8.14%	5.20%	-12.15%
$B_k \leq 7500$	5.48%	**5.01%**	-4.26%	-0.31%	-8.77%	5.44%	-15.03%
$B_k \leq 10000$	**5.52%**	5.00%	-4.49%	-0.33%	**-8.98%**	**5.51%**	**-15.04%**
`AdaptivePacing-Click`							
low-click	**3.39%**	**3.15%**	**-2.45%**	**-0.16%**	**-5.48%**	**3.30%**	-7.92%
high-click	3.13%	2.65%	-2.98%	-0.31%	-5.30%	2.93%	**-10.98%**
`AdaptivePacing-Time`							
$t \leq 1140$	6.11%	5.42%	**-4.19%**	**-0.48%**	-8.76%	5.89%	**-22.65%**
$t \leq 1200$	6.19%	5.50%	-4.54%	**-0.48%**	-9.07%	5.97%	-21.87%
$t \leq 1260$	**6.24%**	**5.52%**	-4.86%	-0.50%	**-9.42%**	**6.02%**	-20.96%
`AdaptivePacing-BudgetSpent`							
PS $\geq$ 50%	6.46%	5.62%	-3.85%	-0.63%	**-8.88%**	6.25%	-19.66%
PS $\geq$ 75%	6.44%	5.66%	-2.90%	-0.57%	-8.07%	6.23%	**-20.73%**
PS = 100%	**6.48%**	**5.73%**	**-1.96%**	**-0.54%**	-7.27%	**6.27%**	-20.42%

For the platform, the information that can aid in distinguishing different campaigns comprises of: (1) the total budget of a campaign B_k; (2) its maximum bid $v_{k,i}$ for each impression; (3) the current time t, (4) a campaign's remaining budget at round t. Given these, we first designed and experimented on three simple heuristics that restrict `AdaptivePacing` (Algorithm 1) to some campaigns at certain time, while not adopting budget pacing for others:

(1) **`AdaptivePacing-Budget`**: Within eBay's sponsored search program, many high-budget campaigns don't exhaust their budget even under no-pacing. Hence, pacing these campaigns would only hinder them from bidding on impressions despite their capability to spend more. Observing this, our first heuristic, `AdaptivePacing-Budget`, would only apply `AdaptivePacing` to campaigns with initial budgets below a budget threshold $\bar{B}$. Initial trials with various threshold levels $\bar{B} = 5000, 7500, 10000$ showed increased spending when `AdaptivePacing` was not universally applied, but also resulted in fewer impressions and clicks, along with a higher pacing error, indicating a trade-off when pacing high-budget campaigns (see Table 3).

(2) **AdaptivePacing-Click**: Another way to differentiate the campaigns is to examine their *click opportunities*, defined as $C_k = B_k / \max_i v_{i,k}$, which is the minimum number of clicks they can generate if depleting their existing budget. We tried only pacing the low-click campaigns with bottom 50% click opportunities, and high-click campaigns with top 50% click opportunities respectively (see Table 3). In both cases, we have observed an increase in spending, yet again accompanied by a trade-off in impressions, clicks, and pacing error when compared to applying AdaptivePacing universally. Interestingly, pacing the low-click campaigns resulted in better performance, suggesting that pacing is more advantageous for campaigns with limited click opportunities.

(3) **AdaptivePacing-Time**: Finally, we consider a heuristic that paces in earlier time periods and halts pacing for all campaigns later on. Given that our simulation is run on a daily basis, we tried only applying AdaptivePacing to all campaigns before 7PM, 8PM and 9PM (i.e., $t \leq 1140, 1200, 1260$); see Table 3. We observe a slight increase in spending compared to performing AdaptivePacing all day. Interestingly, the number of impressions/clicks doesn't drastically decline, and there is even a noticeable enhancement in pacing error. The efficacy of AdaptivePacing-Time primarily stems from the fact that our simulation, similar to many real-world systems, are conducted on a daily/weekly basis and any remaining budget is forfeited by the end of the time horizon. On the contrary, the theoretical guarantees tied to AdaptivePacing are established asymptotically, and hence become less informative in a real-world system.

4.1.4 *What are the right campaigns to pace?*

Our simple heuristics in Section 4.1.3 have shed some light on the outcomes from differentiating treatments for different advertisers at different times. Given the promising theoretical implications of the optimization-based framework that motivates the design of AdaptivePacing, we are also prompted to consider which campaigns would/would not benefit from bid pacing from a theoretical perspective.

To start, we first consult the theoretical results in [5], which suggest that under ideal assumptions and the asymptotic time horizon, AdaptivePacing would lead the pacing multipliers $\mu_{k,t}$ of campaign k to approximately converge to some "optimal" pacing multiplier $\mu_k^\star$. For campaigns whose budget constraints are not tight under the optimization framework in (3), their optimal pacing multiplier would simply be $\mu_k^\star = 0$; while for campaigns with tight budget constraints, there exist a positive multiplier $\mu_k^\star > 0$ that they should adopt in hindsight. (See Theorem 4.3 in [5].)

Even though the results above may not align perfectly with real-world scenarios where some crucial assumptions fail to hold, we can still derive valuable insights. At a high level, these observations emphasize the significance of distinguishing between campaigns with binding/non-binding budget constraints when addressing (3). In light of this, if the goal of AdaptivePacing is to help campaigns identify their optimal pacing multiplier over time, one strategy is to only apply AdaptivePacing to campaigns anticipated to have binding budget constraints, which should expedite the process.

Our heuristic, AdaptivePacing-BudgetSpent, is motivated from this idea—we first consider the proportion of budget that each campaign spent under no-pacing from prior data (denoted by PS), and restrict pacing to campaigns that have spent a sufficient proportion of their budget. Table 3 records the attained metrics when AdaptivePacing is only applied to campaigns that spent over 50%, 75% and 100% of their initial budget. Surprisingly, by only pacing the campaigns that completely depleted their budget (which typically only accounts for a very small fraction of all campaigns in sponsored search programs), we in fact attain the best overall metric compared to prior methods. Compared to applying AdaptivePacing to all campaigns (see Table 1), only pacing the campaigns that need pacing the most would maintain the increase in terms of impressions/clicks, while ensuring that the spending does not decrease too much, satisfying the platform's business needs.

We conjecture that our simple heuristics AdaptivePacing-Budget and AdaptivePacing-Click, discussed in Section 4.1.3, improve our metrics precisely because the total budgets and/or maximum attainable number of clicks are rough indicators of whether a campaign has binding budget constraints or not. Our heuristic AdaptivePacing-Time, on the other hand, leads to improvements essentially because it helps all campaigns whose budget constraints are not binding reach their optimal multiplier 0 instantaneously. Nonetheless, it is also evident that these heuristics are far from perfect indicators for whether the budget constraints are binding for each campaign, and can lead to a tradeoff in our metrics.

Our promising result for AdaptivePacing-BudgetSpent suggest that a platform can look for ways to predict the likelihood of a campaign depleting their budget if no pacing were adopted. One simplistic approach is to perform no-pacing for all campaigns for a short period (e.g., one week), and identify the campaigns that have nearly depleted their budget. The platform can then rely on such information to restrict AdaptivePacing to the budget-binding campaigns that need pacing the most.

4.2 Comparison and connection with other budget pacing approaches

In practice, a number of budget pacing methods have been proposed and adopted in search advertising programs. In this section, we address the following question: **How does the optimization-based AdaptivePacing method measure up against other commonly adopted budget pacing methods?** In Section 4.2.1, we focus our attention on the PID controller, which is a heuristic widely adopted in practice. In Appendix A, we further offer an extended discussion that compare AdaptivePacing against throttling approaches.

4.2.1 **AdaptivePacing** *versus PID controllers.*

The PID controllers are widely used in the industry as heuristics for budget pacing [20, 21]. However, there in fact exists a somewhat surprising connection between AdaptivePacing and the PID controller.

Very recently, the authors of [6] connects the optimization framework (3), which motivated AdaptivePacing, and the PID controller approaches. At a high level, the AdaptivePacing method can be viewed as a special case of a dual-based PID controller. To see that, if we use a PID controller to update the pacing multiplier μ_t, the update would work as follows:

$$\mu_{t+1} = \max\left(\mu_t - \lambda_P g_t - \lambda_I \sum_{s=0}^{t-1} g_{t-s} - \lambda_D \left(g_t - g_{t-1}\right), 0\right) \quad (7)$$

where $\lambda_P, \lambda_I, \lambda_D > 0$ are respectively the step sizes associated with the proportional (P), integral (I) and derivative (D) terms. Under the optimization framework in (3), the error term is exactly our subgradient $g_t = \rho_{k,t} B_k - \tilde{z}_{k,t}$. Given the form in (7), `AdaptivePacing` essentially works as a special case of the PID controller, also known as a P controller, where $\lambda_I = \lambda_D = 0$. The PID controller additionally incorporates (i) past *momentum* via the integral term, which has been shown effective in leading to faster convergence in optimization-based methods [16] and (ii) *optimism* via the derivative term, which can improve the performance of the gradient descent [11].

In view of the close connection between `AdaptivePacing` and the PID controller, we consider a dual-based PID controller as a special extension of `AdaptivePacing` with practical importance, and test the performance of PID controller in our test bed. In practice, the PID controllers are often designed to update a control variable (in our case, $\mu_{k,t}$), and the control variable then directly impacts the bid factor defined using some function f. That is, for $i \in \mathcal{I}_t$, the campaign k posts bid of the following form $b_{i,t} = f(\mu_{k,t}) \cdot v_{i,t}$. Note that here, under the PID controller defined in (3), our control variable $\mu_{k,t}$ would increase if our realized spend $\tilde{z}_{k,t}$ at round t exceeds our ideal budget $\rho_{k,t} B_k$, and decrease vice versa. Hence, f should be a decreasing function such that the campaign would decrease its bid if it overspends, and increase its bid otherwise.

In reality, a number of choices of f can be potentially adopted. In Table 4 we test the performance of the PID controller under a variety of definitions of the bid factor $f(\mu)$. We observe that different forms of the bid factor $f(\mu)$ would prioritize different metrics (for example, $\exp(-\mu)$ appears to be the best in terms of reducing pacing error, while $1/(1+\mu)$, the bid factor we adopt in `AdaptivePacing`, performs best in increasing number of impressions and clicks and surface rate as well as reducing cost-per-click for advertisers.

$f(\mu)$	Imps	Clks	Rev	CTR	CPC	SR	PE
$-1/5\mu + 1$	4.31%	4.03%	**-1.43%**	**-0.29%**	-5.17%	2.47%	-32.07%
$\exp(-\mu)$	5.47%	4.55%	-2.31%	-0.81%	-6.49%	4.66%	**-38.17%**
$1/(1+\mu)$	**5.99%**	**5.52%**	-1.85%	-0.33%	**-7.05%**	**5.84%**	-22.28%

Table 4: Performance of dual-based PID controllers.

PID controllers, despite their wide adoptions, often lack clarity in choices of their bid functions and control variable updates. Therefore, establishing the connection between the PID controller approach and `AdaptivePacing` is valuable in that it allows us to extend all of the theroetical insights established for `AdaptivePacing` and its optimization framework to PID controller-based approaches.

4.2.2 ***Comparison of Budget Pacing Approaches under different spending curves.*** We conclude our experiment section by comparing the spending curves achieved by different types of budget pacing approaches.

Recall from our discussion in Sections 2.3 and 4.1.1 that by tuning the target spending curve, captured by $\rho_{k,t}$, the campaigns that adopt `AdaptivePacing` can choose the target spending curve best suited to their need, such as the traffic curve or the uniform spending curve. Given our discussion in Section 4.2.1, since the PID controller can be considered as an extension to the `AdaptivePacing` method, it can also adapt to different target expenditure curves

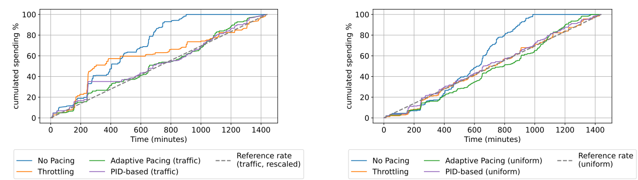


Figure 1: Spending curve of the top campaign under different budget pacing methods. The reference rate has been rescaled according to the target spending curve (left: traffic curve; right: uniform spending curve).

(recall from (7) that the error term g_t directly depends on $\rho_{k,t}$). On the contrary, the `Throttling` method from [19], primarily targets the uniform spending curve and does not have a mechanism that adapts to various spending curves.

In Figure 1, we compare the spending patterns for the top campaign (the campaign with the highest number of clicks), under no-pacing, `AdaptivePacing`, `Throttling` and the PID controller respectively. We also let `AdaptivePacing` and the PID controller adopt two different target spending curves—the traffic curve and the uniform spending curve. Under no-pacing, this campaign would deplete its budget early in the horizon and miss out on all opportunities later on. We observe that since `Throttling` mainly encourages uniform spending, it performs well in helping the campaign achieve uniform spending, yet cannot adapt well when the campaign would like its spending to follow the traffic curve. Both `AdaptivePacing` and the dual-based PID controller (which is an extension of `AdaptivePacing`), on the other hand, shows outstanding performance in aligning the spending pattern of the top campaign to its target, whether the target spending curve is the traffic curve or the uniform spending curve. Given that the PID controller extends directly from `AdaptivePacing`, it usually aligns slight more closely with the target spending curve compared to `AdaptivePacing`. That being said, the performance of `AdaptivePacing` is already excellent as seen from Figure 1, which again validates the efficacy of `AdaptivePacing` in our real-world system.

5 CONCLUSIONS AND FUTURE DIRECTIONS

In this study, we investigate the effectiveness of an optimization-based budget pacing strategy, and its many variants, in the context of eBay's sponsored search environment. By leveraging an optimization framework, we are able to integrate theoretical insights and enhance the bid shading methodology, thereby aligning it with both the commercial objectives of the platform and the specific goals of advertisers and their campaigns. The transition from heuristic-driven to theory-informed budget pacing practices holds promise for transforming operational strategies across various sectors, fostering a more robust and equitable advertising ecosystem.

For future research, there are a few promising directions to pursue. We could incorporate additional business-centric or user-centric elements into our model, including return on investment (ROI) considerations and long-term customer value constraints. Another prospective avenue is the integration of cross-channel marketing effects, where budget pacing can be optimized not only within a single sponsored search program but also across multiple advertising channels for holistic campaign performance.

REFERENCES

[1] Deepak Agarwal, Souvik Ghosh, Kai Wei, and Siyu You. 2014. Budget pacing for targeted online advertisements at linkedin. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1613–1619.

[2] Santiago Balseiro, Anthony Kim, Mohammad Mahdian, and Vahab Mirrokni. 2017. Budget management strategies in repeated auctions. In *Proceedings of the 26th International Conference on World Wide Web*. 15–23.

[3] Santiago R Balseiro, Omar Besbes, and Gabriel Y Weintraub. 2015. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science* 61, 4 (2015), 864–884.

[4] Santiago R Balseiro, Omar Besbes, and Gabriel Y Weintraub. 2019. Dynamic mechanism design with budget-constrained buyers under limited commitment. *Operations Research* 67, 3 (2019), 711–730.

[5] Santiago R Balseiro and Yonatan Gur. 2019. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science* 65, 9 (2019), 3952–3968.

[6] Santiago R Balseiro, Haihao Lu, Vahab Mirrokni, and Balasubramanian Sivan. 2022. From Online Optimization to PID Controllers: Mirror Descent with Momentum. *arXiv preprint arXiv:2202.06152* (2022).

[7] Denis Charles, Deeparnab Chakrabarty, Max Chickering, Nikhil R Devanur, and Lei Wang. 2013. Budget smoothing for internet ad auctions: a game theoretic approach. In *Proceedings of the fourteenth ACM conference on Electronic commerce*. 163–180.

[8] Zhaohua Chen, Chang Wang, Qian Wang, Yuqi Pan, Zhuming Shi, Zheng Cai, Yukun Ren, Zhihua Zhu, and Xiaotie Deng. 2022. Dynamic budget throttling in repeated second-price auctions. *arXiv preprint arXiv:2207.04690* (2022).

[9] Dragos Florin Ciocan and Krishnamurthy Iyer. 2021. Tractable equilibria in sponsored search with endogenous budgets. *Operations research* 69, 1 (2021), 227–244.

[10] Vincent Conitzer, Christian Kroer, Eric Sodomka, and Nicolas E Stier-Moses. 2022. Multiplicative pacing equilibria in auction markets. *Operations Research* 70, 2 (2022), 963–989.

[11] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. 2017. Training gans with optimism. *arXiv preprint arXiv:1711.00141* (2017).

[12] Jason Gaitonde, Yingkai Li, Bar Light, Brendan Lucier, and Aleksandrs Slivkins. 2022. Budget pacing in repeated auctions: Regret and efficiency without convergence. *arXiv preprint arXiv:2205.08674* (2022).

[13] Gagan Goel, Vahab Mirrokni, and Renato Paes Leme. 2020. Clinching auctions with online supply. *Games and Economic Behavior* 123 (2020), 342–358.

[14] Chinmay Karande, Aranyak Mehta, and Ramakrishnan Srikant. 2013. Optimizing budget constrained spend in search advertising. In *Proceedings of the sixth ACM international conference on Web search and data mining*. 697–706.

[15] Niklas Karlsson and Jianlong Zhang. 2013. Applications of feedback control in online advertising. In *2013 American control conference*. IEEE, 6008–6013.

[16] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[17] Kuang-Chih Lee, Ali Jalali, and Ali Dasdan. 2013. Real time bid optimization with smooth budget delivery in online advertising. In *Proceedings of the seventh international workshop on data mining for online advertising*. 1–9.

[18] Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. 2007. Adwords and generalized online matching. *Journal of the ACM (JACM)* 54, 5 (2007), 22–es.

[19] Phuong Ha Nguyen, Djordje Gligorijevic, Arnab Borah, Gajanan Adalinge, and Abraham Bagherjeiran. 2023. Practical Budget Pacing Algorithms And Simulation Test Bed For EBay Marketplace Sponsored Search. In *AdKDD Workshop 2023 at the 29th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (AdKDD 2023)*.

[20] Xun Yang, Yasong Li, Hao Wang, Di Wu, Qing Tan, Jian Xu, and Kun Gai. 2019. Bid optimization by multivariable control in display advertising. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 1966–1974.

[21] Weinan Zhang, Yifei Rong, Jun Wang, Tianchi Zhu, and Xiaofan Wang. 2016. Feedback control of real-time display advertising. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. 407–416.

[22] Weinan Zhang, Shuai Yuan, and Jun Wang. 2014. Optimal real-time bidding for display advertising. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1077–1086.

A COMPARISON WITH THROTTLING

As discussed in Section 1, budget pacing approaches are generally categorized as either *throttling*, which allows campaigns to join auctions probabilistically, or *bid shading*, which modifies the bid value for ad allocation. In practice, most budget pacing methods can be implemented in either manner. Throttling, often considered more favorable for emerging ad platforms, is a cruder method that supports system triage. Moreover, advertisers have shown a preference for throttling over bid-altering approaches in sponsored search [14]. Previous research [19] has demonstrated the system-level benefits of various throttling approaches in the context of eBay sponsored search, albeit without extensive theoretical backing.

Throttling approaches can further be categorized based on the information they utilize, which includes budget reset at different times (strategically resetting the budget to maximize competition), remaining budget-based approaches [18], remaining budget and time-based approaches [14], and remaining ad impressions and clicks-based approaches [14]. Throttling-based budget pacing methods often compromise the impressions of ad campaigns to enhance overall competition, often achieved by reducing impressions, especially earlier in the horizon. However, there are minimal guarantees regarding the availability or quality of future ad opportunities.

In this section, we consider two different throttling approaches: (1) `Throttling`: the throttling approach based on remaining budget and time introduced in [14], which is shown to be the most effective among a number of throttling approaches considered by [19], specifically in the test bed of eBay sponsored search; (2) `AdaptiveThrottling`: a throttling method that directly uses the bid multiplier we establish for `AdaptivePacing`, i.e., $1/(1+\mu_{k,t})$, as the probability that a campaign would join an auction.

We evaluate the performance of the two throttling approaches via the test bed described in Section 3.1, and compare the metrics attained by throttling against those of `AdaptivePacing` in Table 5 below. From Table 5, we observe that `AdaptivePacing` outcompetes `Throttling` and `AdaptiveThrottling` in almost all of the key business metrics, such as the number of impressions/clicks, surface rate, CPC, etc. (See, also, Appendix B for illustrations that compare `AdaptivePacing` with `Throttling`.) The only metric where throttling based approaches excel is the pacing error, which can gets reduced by as much as 50%. However, as we discussed in Section 4.1.1, this is achieved at the expense of other vital metrics. For instance, both throttling methods would lead a decrease in the number of impressions/clicks and the surface rate even compared to no-pacing, which can be less favorable to both the platform and the advertisers.

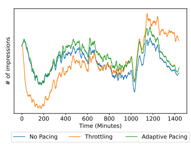
It is also noteworthy that that the bid multiplier $1/(1+\mu_{k,t})$ in `AdaptivePacing` fails to yield satisfying performance when it is used as a throttling threshold in `AdaptiveThrottling`. Such result aligns with previous findings in [1, 19].

Table 5: Performance of `AdaptivePacing`, `Throttling` and `AdaptiveThrottling`.

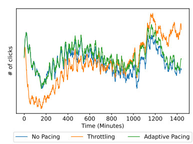
Method	Imps	Clks	Rev	CTR	CPC	SR	PE
`AdaptivePacing`	**6.21%**	**5.48%**	-5.39%	**-0.50%**	**-10.20%**	**6.01%**	-18.72%
`Throttling`	-0.45%	-1.10%	**-1.89%**	-0.88%	-0.28%	-7.76%	**-51.62%**
`AdaptiveThrottling`	-0.28%	-2.63%	-3.11%	-2.14%	-0.08%	-9.14%	-20.96%

B ILLUSTRATIONS OF BUSINESS METRICS

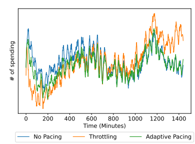
In this section, we illustrate how `AdaptivePacing` impacts our key business metrics, stated in Section 3.2. In Figure 2, we plotted the evolution of number of impressions (**Imps**), number of clicks (**Clks**), total ad revenue (**Rev**), the average click-through rate (**CTR**), the average cost-per-click (**CPC**), and the surface rate (**SR**) of campaigns in the sponsored search program throughout the day. We compare the business metrics under no-pacing, `Throttling` (see Appendix A and [19] for a description of this throttling method) and `AdaptivePacing`. Note that to preserve sensitive information, all of the plots presented in this section have been detrended and the numerical values have been omitted.



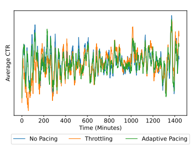
(a) Imps.



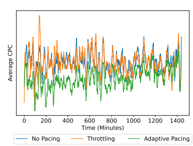
(b) Clks.



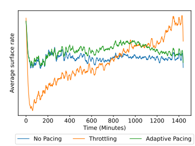
(c) Rev.



(d) CTR.



(e) CPC.



(f) SR.

Figure 2: Evolution of business metrics under no pacing, `Throttling` and `AdaptivePacing`. All plots are detrended.

Figure 2 illustrates the distinct strategies of `Throttling` and `AdaptivePacing` for budget pacing. `Throttling` paces budgets by reducing a campaign's probability of entering auctions early in the horizon, conserving funds for later use. Conversely, `AdaptivePacing` employs bid shading to moderate bid amounts based on a campaign's spending pace, rather than restricting auction participation. This method ensures campaigns can capitalize on available spending opportunities. This contrast in strategies is highlighted in Figures 2a, 2b, and 2f, where `Throttling` appears to limit early-day exposure to reserve funds for potential impressions or clicks later in the day. In contrast, `AdaptivePacing` provides a more balanced spending approach throughout the day. It is surprising to see that in terms of key metrics such as number of impressions/clicks and the surface rate, `AdaptivePacing` strictly dominates no-pacing throughout the horizon.

Further, Figure 2e reveals that while `Throttling` has a negligible impact on cost-per-click (CPC), `AdaptivePacing` effectively lowers the CPC through its bid shading mechanism in a second-price auction setting. This reduction in CPC enables campaigns to secure more impressions or clicks and simultaneously increases the advertisers' net utility per click.

All of the experiment results above reinforce the premise that bid shading approaches like `AdaptivePacing`, as suggested by its optimization-based framework, contributes significantly to maximizing advertisers' utility. These complement the theoretical results established previously in a similar vein (see Theorem 4.3 in [2]).