User Interaction with Online Advertisements: Temporal Modeling and Optimization of Ads Placement

(Article begins on next page)

19 April 2024

User Interaction with Online Advertisements: Temporal Modeling and Optimization of Ads Placement

LUCA VASSIO, Politecnico di Torino, Italy
MICHELE GARETTO, Università di Torino, Italy
CARLA CHIASSERINI, Politecnico di Torino, Italy
EMILIO LEONARDI, Politecnico di Torino, Italy

We consider an online advertisement system and focus on the impact of user interaction and response to targeted advertising campaigns. We analytically model the system dynamics accounting for the user behavior and devise strategies to maximize a relevant metric called click-through-intensity (CTI), defined as the number of clicks per time unit. With respect to the traditional click-through-rate (CTR) metric, CTI better captures the success of advertisements for services that the users may access several times, making multiple purchases or subscriptions. Examples include advertising of on-line games or airplane tickets. The model we develop is validated through traces of real advertising systems and allows us to optimize CTI under different scenarios depending on the nature of ad delivery and of the information available at the system. Experimental results show that our approach can increase the revenue of an ad campaign, even when user's behavior can only be estimated.

CCS Concepts: • **Information systems** → **Social advertising**; *Social recommendation*; *Personalization*; **Display advertising**; **Social advertising**; *Social recommendation*; *Personalization*; *Social recommendation*; *Social recommendation*; *Personalization*;

Additional Key Words and Phrases: Online advertisements, Recommendation systems, Ads placement, CTR, User behaviour

1 INTRODUCTION

Online advertising is a market estimated in 2018 at more than 100 billion dollars, in the USA alone [3]. This success has largely arisen with the usage of automatic processes that enable highly targeted advertising. In this paper, we consider an online system for targeted advertising, as illustrated in Fig. 1. On one side, a Publisher provides an Ad Server with a stream of available slots, i.e., portions of the user's navigation experience where ads can be inserted. On the other side, an Advertiser provides a stream of ads that can potentially fill in those slots. The Advertiser will manage the slots according to its advertisement campaign. A subset of the available slots is matched with ads

Authors' addresses: Luca Vassio, Politecnico di Torino, Italy, luca.vassio@polito.it; Michele Garetto, Università di Torino, Italy, michele.garetto@unito.it; Carla Chiasserini, Politecnico di Torino, Italy, carla.chiasserini@polito.it; Emilio Leonardi, Politecnico di Torino, Italy, emilio.leonardi@polito.it.

and generates a stream of impressions shown to each given user, who might decide to perform actions on them, such as clicks. Clicks can further turn into subscriptions to a service or purchases and eventually money for the advertiser.[1] The user can be seen as a filtering funnel, where just a subset of the delivered impressions becomes clicks/actions, and even less becomes purchases. The intelligence of the system resides in the Ad Server, which decides which impressions to deliver to each specific user, and at which time instants, usually guided by analytics collected on the user itself.
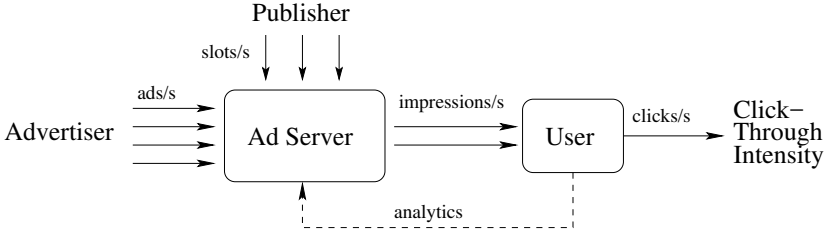

Fig. 1. High-level view of the studied advertising system

Both the Publisher and Advertiser's revenues grow with the increase of the users' actions over the duration of the campaign. This is particularly true for advertisements campaigns for services that the users may access several times over a given period, making multiple purchases and generating a stream of revenue. Examples include advertising of airplane tickets or pay-to-play online games. In the latter case, users pay the game each time they play, and therefore they are pushed to resume playing through advertisements. In their attempt to maximize their revenues, Publisher and Advertiser should take into account that the likelihood of a specific user to perform a valuable action upon an impression may be heavily impacted by the history of shown impressions. In other words, the number and temporal spacing of impressions can have a profound impact, as it has already been recognized [8–10, 14, 19]. A user overwhelmed with impressions arriving too close in time might be less likely to perform actions on them, because he/she get annoyed by the ads.

A standard metric to measure the quality of online advertisements is the action-through-rate, defined as the number of actions divided by the number of impressions [17]. When the considered actions are the clicks made on the ads, the metric becomes the well-known click-through-rate (CTR). The CTR does not take into account the temporal spacing by which impressions are shown to the user. For this reasons, in our work we consider the click-through-intensity (CTI) metric, expressed in number of clicks per time unit. To this end, we study the detailed temporal dynamics of the advertising system in Fig. 1 by developing a model that incorporates the user's reaction. Then we estimate the likelihood that the user will perform a valuable action on a particular impression shown at a given time.

Our main contributions are as follows.

- We introduce a stochastic framework for the interaction of online advertising system with the user, supported by evidences on traces from real advertising systems. To our knowledge, this is the first analytical model capturing the main features of the above system;
- We identify different regimes of the proposed system and devise strategies to optimize the CTI metric and the frequency capping of an ad campaign. Such strategies can provide useful theoretical benchmarks for the deployment of better advertising platforms.

---

[1]For brevity, hereinafter we will refer to purchases only as a possible consequence of a click.

The rest of the paper is organized as follows. Sec. 2 discusses related work and emphasizes the novelty of our study. Sec. 3 introduces the system model and the CTI performance metric, and it shows the need for maximizing CTI rather than the traditional CTR metric. The strategies we envision to optimize the system performance are presented in Sec. 4, along with numerical results. Sec. 5 describes a methodology to estimate the user behavior in realistic cases where the Ad Server does not have full knowledge of it. Finally, Sec. 6 draws conclusions and discusses possible directions for future research.

2 RELATED WORK

Automated optimization of online ad placement, targeting, and bid prices, have being subject of extensive investigation. Although very aggressive ways to display ads to the users are often adopted, the success of a campaign depends also on the quality of user engagement. In particular, overly intrusive ads are often completely avoided by viewers, and tend to have a very detrimental effect on user experience [18, 28]. Purchase events may be driven by habituation and boredom [6]. The first one characterizes the inertia of buying the same product or service over time, while the latter makes the user tired and seek for variety. Such tendencies can coexist within the same consumer, evolving over time.

Nowadays, behavioral and contextual targeting have emerged as techniques to increase the efficiency and profits of digital advertisement. In [12] authors analyze the economic implications of behavioral targeting, showing that the revenue for the publisher can double by using this technique. Several empirical studies have investigated the benefits of behavioral targeting, focusing on traditional CTRs [24, 34]. The competitive interaction among content publishers over a social network, and the actions that can be taken to increase content visibility, have been studied through the lens of game theory in [5].

Similar human factors have been considered in the related field of recommendation systems based on collaborative filtering [16]. In addition to accuracy, several authors have suggested various ways to combine other metrics such as diversity, coverage, and serendipity [11, 13, 25, 26]. There have been also attempts to incorporate temporal dynamics of user behavior in techniques for recommendation systems. In particular, time-decaying weight functions are usually introduced to capture the fact that more recent data better reflect a user's current preference [15, 22]. However, tracking the temporal dynamics of customer preferences is generally recognized as a challenging problem [21], due to the existence of several possible interest-drift patterns [10].

The majority of existing models ignore the fact that users may get annoyed or bored of recommendations, despite their past interactions. One exception is the work in [20], where authors show that user's temporal consumption of familiar items is driven by boredom. They propose a semi-Markov model with two latent psychological states, sensitization and boredom, to characterize the user revisit times to an item. The authors of [14] also introduce boredom to explain the cyclic pattern of individual choices as well as social trends. They propose a model in which boredom is proportional to the total accumulated memory for an item, where each consumption of the item adds a term that drops geometrically over time. The analysis in [23] accounts for the temporal evolution of user interests through attraction or aversion towards past suggestions. In [19] authors propose a model of user response to an ad campaign as a function of both the interest match and the past exposure.

To contain the negative effects of user boredom/annoyance, many existing ad serving technologies already offer to advertisers a configurable option called frequency capping. For example, the popular platform Google AdWords (now Google Ads) allows setting a limit to the number of impressions an individual user will see per day, per week, or per month [1]. Although frequency capping is a standard practice, there are no well established methodologies to set the threshold, and most

advertisers resort to trial-and-error or rough guidelines, e.g., 3 views/visitor/day. In [30] the authors propose to set frequency capping policies for different online marketing segments using Markov decision processes with various features such as CTR of category, hour of day and CTR within the last 24 hours. To the best of our knowledge, *we are the first to theoretically investigate the problem of dynamically optimizing over time the frequency of an ad campaign using a behavioral model based on received impressions.*

3 SYSTEM MODEL

First, we describe the metric we seek to optimize (Sec. 3.1), i.e., CTI. Then we propose a model on how a user is affected by received impressions, and how likely the user will click on a given impression (Sec. 3.2). Then, we will show the applicability of the response function model on real traces (Sec. 3.3). Then we explain the different scenarios that arise depending on how impressions can be delivered to the user by the Ad Server (Sec. 3.4). At last, we introduce a baseline strategy to propose advertisements to users (SLT), showing the need to optimize CTI to maximize the revenue of an ad campaign (Sec. 3.5).

3.1 Click-through-intensity metric

We assume that the Advertiser is paying to the Publisher a constant cost-per-action or cost-per-click. On the other hand, the Advertiser gets a revenue proportional to the number of actions performed by users. Under this model, both Advertiser and Publisher are interested in increasing the number of actions performed by each user in a certain (usually long) period of time. We call this metric, directly proportional to profits of both Advertiser and Publisher, click-through-intensity (CTI).

Considering a time interval from $t = t_{start}$ to $t = t_{end}$ and impressions i displayed to the user at times t_i, the empirical CTI for this user is defined as:

$$CTI = \frac{\sum_{i:t_{start} \leq t_i \leq t_{end}} \mathbf{1}_{click(i)}}{t_{end} - t_{start}}$$

where $\mathbf{1}$ is the indicator function, and *click(i)* is the event that the user performs a click on impression i. The probability of performing a click depends on the history of past impressions and, as detailed later, it depends on the internal state of the user. In a nutshell, CTI measures the rate of performed clicks. Our objective will be to maximize CTI under different ergodic regimes, i.e., assuming that t_{end} tends to infinity[2]. This is different from maximizing the traditional CTR: indeed, CTR measures the effectiveness of ads, and a higher CTR does not imply a larger number of clicks (e.g., if the rate of clicks largely decreases). CTR is a very important metric, but it does not capture well the impact of impression frequency in the case of repetitive actions on the same campaign, e.g., access to a website or purchase of a service.

3.2 Model of user excitation and response

We associate to each user a time-varying, non-negative real value $U(t)$ that we call *user excitation*. The user excitation keeps track of the cumulative effect of the impressions shown to the user, assuming that the impact (deterministic or random) of each impression decays over time. As usually done in many physical systems, and supported by the Ebbinghaus's model [27], the impact on the user of each impression is assumed to decay exponentially over time, with parameter α, representing in our context the physiological process of forgetting. Parameter α can be inferred for the specific user, or just estimated from the average behavior of a larger set of users. We assume, for now, that

---

[2]We assume that the advertising campaign is long enough to neglect effects due to its finite duration.

the exact value of α is known to the system. In Sec. 5.1, we evaluate the impact of an erroneous estimate of α.

When an impression i arrives, it increments the excitation by L_i. In the simplest case, L_i is just a constant, but for greater generality we allow it to be a random variable accounting for various effects related to the user behavior, the environment, etc. For simplicity, we will assume that L_i are i.i.d. positive random variables with the same distribution and that the distribution of L_i has all polynomial moments finite. Without lack of generality, we can assume that user excitation has mean equal to 1, i.e., $\mathbb{E}[L_i] = 1$.

Fig. 2 shows an example of the temporal evolution of $U(t)$ in the case of a user receiving a sequence of four impressions displayed at times t_i, $i = 1 \ldots 4$.
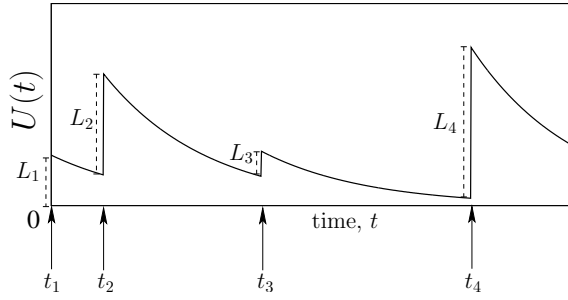

Fig. 2. Example of evolution of user excitation $U(t)$.

By denoting with t the current time instant and with t_i the time at which the user has received the i-th previous impression, the current value of user excitation is given by:

$$U(t) = \sum_{i:t_i<t} L_i e^{-\alpha(t-t_i)} . \tag{1}$$

We will consider $U(t)$ to be left continuous. Two different cases for the stimuli L_i may take place:

- **Perfect excitation information.** The Ad Server knows the times t_i and the values L_i. Under perfect information on L_i and α, the system has full knowledge of the instantaneous user excitation $U(t)$. This is an analytically tractable ideal case, in which we can characterize the optimal strategy, thus deriving upper bounds to the system performance.
- **Partial excitation information.** In this more realistic case, the Ad Server, in addition to times t_i, has inferred a probability distribution over L_i, thanks, for example, to detailed tracking of the interaction between the user and the ads displayed.

In our stochastic framework, the user reacts to an impression depending on the history of past impressions. We introduce the probability P_a that the user performs an action. To emphasize the dependency of such probability on the user excitation and on the impression just received, i.e., the value L_i., we will denote this probability on impression i displayed at time t_i by $P_a(U(t_i) + L_i)$. We will refer to $P_a(x) : \mathbb{R}^+ \to [0, 1]$ as the user response function, and again assume that this function is known to the system, having been either inferred for the specific user or estimated across a larger set of users. We do not pose particular restrictions to function P_a, e.g., it does not even have to be continuous. However, to avoid trivialities, we always assume that $P_a(x) \to 0$ for $x \to \infty$.

Table 1. Avazu and Outbrain datasets overview.

	Avazu dataset	Outbrain dataset
Total trace time	10 days	14 days
Total number of users	100 000	100 000
Total number of impressions	11 213 904	2 914 236
Total number of clicks	1 968 027	460 384
Average CTR	0.1755	0.1579
Median number of impressions per user	59.0	26.0
Average number of impressions per user	112.1	29.1
Median intertime between impressions	15 min	< 1 min
Average intertime between impressions	6 h 40 min	4 h 55 min

3.3 Experimental evidences

We now provide qualitative examples of user response function P_a derived from experimental data, using traces of real advertising systems. Information about user behavior are sensitive, hence all data and traces must be correctly anonymized before processing [32]. The Avazu dataset, publicly available over the Kaggle platform [2], reports the click/no-click actions performed by 9 million anonymized mobile users on on-line ads, over 10 days. Most of the users in the original dataset are not active, with just 1 impression shown during the period. Since we are interested in users that are exposed to many ads, we restrict the analysis to the 100 000 most active users, generating about 11 million impressions. Characteristics of the used dataset are reported in Table 1.[3]

Assuming, in the absence of further information, that L_i's are deterministic and equal to 1 we computed the evolution of the shot-noise process $U(t)$ for each user using the impression arrival times trying different values of α. For each value of α, we have then evaluated the corresponding empirical user response function[4], using the click/no-click information in the trace. Fig. 3(a) reports the empirical P_a, in the case of $\alpha = 0.3$, using a uniform binning for the user excitation. We emphasize that, although we display aggregate results for all considered users, each user has been treated independently from the others, by reconstructing his/her specific excitation. The dashed line in the plot shows the best least-square fitting of the experimental data by an exponential function, that we will use later on (Sec. 4.5). Results in Fig. 3(a) reveal a significant correlation between the user response and the user excitation[5], suggesting that our methodology can be effectively employed to model and optimize the system (as explained in Sec. 4).

We repeated the same experiment with the Outbrain dataset, also available on the Kaggle platform [4]. The trace contains click/no-click information on roughly 15 million users over a period of 14 days, though again no-clicks have been significantly subsampled to reduce the trace size. In this dataset, most of the users are exposed only to a single burst of very few ads. Again, we restricted our attention to the 100 000 most active users, accounting for approximately 3 million impressions (see Table 1). Fig. 3(b) shows the empirical P_a obtained again for $\alpha = 0.3$. Interestingly, results are similar to those coming from the Avazu dataset. They suggest that, in both datasets, user response is mainly affected by a boredom effect, producing a monotonically decreasing user response function.

---

[3]While the publicly available trace contains all click events, no-click events have been subsampled by Kaggle to limit the trace size, which explains why the observed CTRs are significantly larger than what usually reported.
[4]To avoid transient effects, we discarded for each user the first $2/\alpha$ hours of the reconstructed $U(t)$ process.
[5]We obtained similar results for other values of α around 0.3, whereas correlation tends to vanish for much smaller (≈ 0.01) or much larger (≈ 10) values.

To make our results reproducible and to foster new research on the topic, we decided to make available to the community the used datasets and the software for analyzing them.[6]
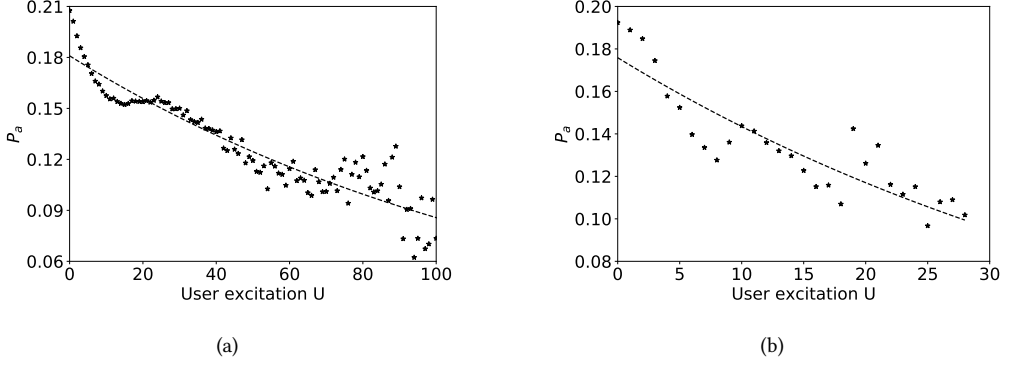
Fig. 3. Experimental P_a obtained from (a) Avazu and (b) Outbrain datasets.

As further evidence of the existence of the boredom effect, in Fig. 4 we show the histogram of the probability that a specific item is clicked at its n-th appearance ($n = 1, \ldots, 5$), given that it has been shown to the same user at least n times and has been clicked by the users. Here we consider the same users as before and we focus on a sequence of specific advertisements. We observe that the considered probability drops sharply with n, with only 12% of clicks at the second impression delivery, and 3% of clicks at the third impression delivery. No item was clicked after its 5-th impression delivery to the same user.

In marketing research, it is also common to consider a non-monotonous user response function with a single peak to jointly account for habituation and boredom. An inverted-U shaped function has been justified by Berlyne's theory of exploratory behavior [7], who first studied the relationship between attractiveness of a stimulus and its familiarity (number of repetitions). Since such a theory

---

[6]Filtered datasets and codes are available at the following public link: https://mplanestore.polito.it:5001/sharing/aSAcHJBtg.

Fig. 4. Probability to click an item at its n-th appearance ($n = 1, \ldots, 5$), given that it has been clicked and shown to the same user at least n times.

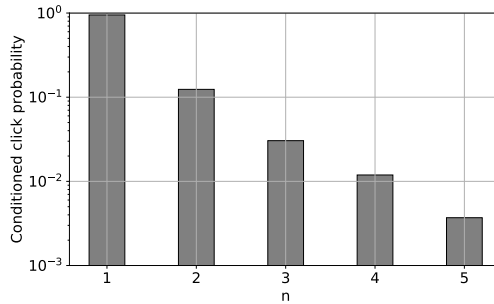has been largely adopted in marketing research, in our evaluation we will also consider a non-monotonous P_a, in addition to a monotonically decreasing function like the simple exponential fitting shown in Figs. 3(a) and 3(b).

At last, we emphasize that, although the way in which we have fitted our model to the Avazu and Outbrain datasets is largely arbitrary, due to intrinsic limitations of the information available in the traces, the main point that we want to make is that *user response and user excitation are indeed significantly correlated*, at least over appropriate time scales. If P_a were independent, or very weakly correlated with U (e.g., a flat function), there would be no room for any optimization, i.e., no reason to introduce any frequency capping: the best strategy would be to overwhelm the users with ads, exploiting all possible impression opportunities.

3.4 Timeliness of ads delivery

Since tracking and profiling technologies permit designing advertising strategies tailored to individuals, we will focus on just one user, whose profile (interests, navigation habits, response history to past advertisements) is known to the system. In the absence of information about the user (e.g., a new visitor), the system will initially use average characteristics of its known users. We will further assume that candidate ads to be sent to the user are qualitatively similar, i.e., they can be considered to be equally interesting to the user according to its current profile information. Note that the Ad Server (see Fig. 1) has to match candidate ads of the Advertiser with available slots of the Publisher. In the following, we refer to a matched pair (ad+slot) as an *impression opportunity*.

For such model, we analyze the following scenarios:

- **Arbitrary delivery.** The system can send impressions to the user at arbitrary time instants. This assumption provides an upper bound to the system performance. Moreover, it can be considered as a good approximation of systems in which there is abundance of impression opportunities, i.e., when we jointly have (see Fig. 1): i) abundance of slots, meaning that the user is online often enough (as compared to the time-scale of the optimal ad delivery rate); ii) abundance of ads, meaning that new ads arrive at sufficiently large rate, and they are delay-tolerant (i.e., they can be delayed, maintaining equal interest to the user).
- **Real-time delivery.** Either because of scarcity of slots, or because of scarcity of ads (not delay-tolerant), the system is forced to select the impressions to send to the user from an online stream of finite rate λ. In particular, we will consider the case in which impression opportunities become available to the system according to a Poisson process of rate λ and the system has to make an instantaneous decision on whether to send each impression to the user or not.
- **Buffer delivery.** This is an intermediate case, which we will explore by simulation, in which ads arrive at the Ad Server at finite rate λ and remain valid for some known time (deterministic or random). Slots are abundant so that ads can be shown to the users at arbitrary times.

Under our assumptions, the profits of an ad-campaign are directly proportional to the CTI, thus our objective will be to maximize such metric under the various scenarios introduced above.

3.5 The baseline SLT strategy

Before diving into the analysis, we propose a baseline frequency capping strategy in the case of partial excitement information and real-time delivery that will be used for comparison in the next sections. We assume impression opportunities arrive according to a Poisson process of rate λ and the Ad Server just delivers the impressions to the user as soon as they arrive. FFig. 5 reports the CTI as a function of the arrival rate λ, assuming $L_i = 1 \ \forall i$, $\alpha = 0.1$, and the monotonically decreasing response function $P_a = 0.1e^{-u}$. Fig. 5 clearly shows that there exists an optimal ad-delivery rate
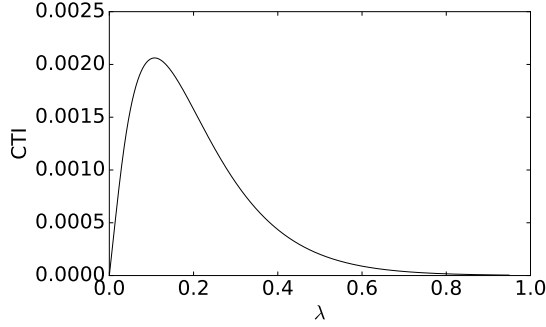
Fig. 5. CTI as a function of delivered impression rate of a Poisson process.

maximizing the CTI. We propose a strategy called stateless thinning (SLT), which performs a thinning of the arriving stream of impression opportunities by sending each opportunity to the user with fixed, independent probability p. In Appendix A we show that the optimal thinning probability p can be analytically derived. Note that, assuming λ large enough, as we vary p the above SLT strategy would precisely achieve the performance shown in Fig. 5 (regarding the x-axis as the rate of thinned impressions). Therefore, an optimal choice of p, possibly by trial and error, would allow us to achieve the maximal value of CTI in Fig. 5.

We argue that this simple SLT strategy, if adopted, could already improve the profits of online campaigns that overwhelm users with uncapped impressions, thus operating on the right portion of the curve shown in Fig. 5. In general the SLT strategy is not optimal, under both arbitrary and real-time delivery scenarios.

We conclude this section with an observation. Under any strictly decreasing user-response (like the one in the example), CTR is maximized under vanishing ad-delivery rate, i.e., by waiting until the user excitation U becomes very small, so that the next opportunity will be accepted with the highest probability. Clearly, so doing one would achieve vanishing CTI, confirming that CTR is not the right metric to consider when maximizing the profit of an ad campaign seeking multiple actions from the user.

4 STRATEGIES WITH PERFECT EXCITEMENT INFORMATION

We start with the case in which the system has full knowledge of the user excitation $U(t)$, i.e., the perfect excitement information case. Let $\{t_i\}_i$ be the sequence of times at which the user is exposed to impression opportunities, i.e., right before the excitation increment due to impression delivery. We observe that $U(t_i)$ is uniquely determined by the triplet $(U(t_{i-1}), L_{i-1}, \tau_i)$, with $\tau_i = t_i - t_{i-1}$. In other words, assuming $U(t_{i-1})$ to be given, $U(t_i)$ is conditionally independent of $\{U(t_j)\}_j$ for $j < i - 1$, whenever τ_i is conditionally independent of $\{\tau_j\}_{j<i}$, $\{U(t_j)\}_{j<i-1}$ $\{L_j\}_{j<i-1}$.

We can thus restrict ourselves to study Markovian policies (i.e., consider policies according to which $\{\tau_i\}_i$ satisfies previous properties), and prove that $\{U(t_i)\}_i$ forms an ergodic Markov process over $\mathbb{R}^+$ (or a compact subset of $\mathbb{R}^+$) under some additional weak assumptions on the distribution of L_i and τ_i.

PROPOSITION 4.1. *Assume that: (i) L_i, which represents the increment in the user excitation upon the delivery of an impression opportunity, exhibits finite polynomial moments, and (ii) τ_i exhibits finite*

polynomial moments and $\tau_i > \delta$ with probability one, for some $\delta > 0$, whenever $U(t_{i-1})$ is sufficiently large. Then $\{U(t_i)\}_i$ represents an ergodic Markov process.

Proof. A proof of the ergodicity of $\{U(t_i)\}_i$ can be given with standard drift arguments. Indeed, $\mathbb{E}[U(t_i) \mid U(t_{i-1})] = \mathbb{E}[(U(t_{i-1}) + L_{i-1})e^{-\alpha\tau_i}] < (\mathbb{E}[U(t_{i-1})] + \mathbb{E}[L_{i-1}])e^{-\alpha\delta} = \mathbb{E}[U(t_{i-1})] - \mathbb{E}[U(t_{i-1})](1 - e^{-\alpha\delta}) + \mathbb{E}[L_{i-1}]e^{-\alpha\delta} < \mathbb{E}[U(t_{i-1})] - 1$ whenever $\mathbb{E}[U(t_{i-1})]$ is sufficiently large. □

Being the process ergodic, we can define $\pi_U(u)$ as the unique stationary distribution of $\{U(t_i)\}$ over a properly defined support. In the following, we study the $\{U(t_i)\}_i$ process under stationary conditions and denote by U the random variable representing the user excitation at the generic time t_i. We also denote by L the generic L_i.

Given the above observations and assumptions, below we derive the optimal Markovian policy under both arbitrary delivery (Sec. 4.1) and real-time delivery (Sec. 4.2) scenarios.

4.1 Arbitrary delivery

Recall that in this scenario the Ad Server can deliver impression opportunities to the user at arbitrary time instants. We will show that in this case the policy that maximizes the CTI is the one that sends a new impression opportunity to the user whenever the excitation $U(t)$ goes back to a fixed, optimal value θ_0, which depends on the response function P_a and the distribution of L_i.

We start by proving the following proposition, which expresses the CTI as the ratio between the average revenue obtained by an impression and the average time between two consecutive opportunities shown to the user.

Proposition 4.2. *The* CTI *achieved by a Markovian policy with stationary distribution* $\pi_U(u)$ *is given by:*

$$\text{CTI} = \frac{\alpha\,\mathbb{E}_U\mathbb{E}_L[P_a(U + L)]}{\mathbb{E}_U\mathbb{E}_L\left[\log\left(1 + \frac{L}{U}\right)\right]} = \frac{\alpha\int_u\int_l P_a(u + l)\,\mathrm{d}F_L(l)\,\mathrm{d}\pi_U(u)}{\int_u\int_l \log\left(1 + \frac{l}{u}\right)\,\mathrm{d}F_L(l)\,\mathrm{d}\pi_U(u)}. \tag{2}$$

Proof. The above result is obtained by applying renewal theory; the proof can be found in Appendix B. □

Looking at (2), it can be seen that the CTI depends on the selected policy only through the stationary distribution $\pi_U(u)$. Therefore, all the Markovian policies with the same stationary distribution lead to the same performance. In particular, we are interested in the optimal policy associated to

$$\pi_U^*(u) = \arg\max_{\pi_U(u)} \frac{\mathbb{E}_U\mathbb{E}_L[P_a(U + L)]}{\mathbb{E}_U\mathbb{E}_L\left[\log\left(1 + \frac{L}{U}\right)\right]}$$

where for now we assume that a maximum exists over $\pi_U(u)$ (its existence is shown below).

Next, within the class of Markovian policies, we define a subclass of policies enforcing $U(t_i) = \theta$, $\forall i$, i.e., policies that expose the user to a new impression opportunity whenever the excitation goes back to a given threshold θ. In the following, we will generally refer to such polices as *threshold-based.* For threshold-based policies, the expression of CTI reduces to:

$$\text{CTI} = \frac{\alpha\int_l P_a(\theta + l)\,\mathrm{d}F_L(l)}{\int_l \log\left(1 + \frac{l}{\theta}\right)\,\mathrm{d}F_L(l)}.$$
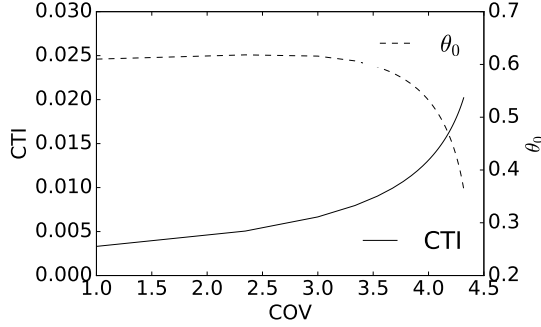
Fig. 6. Optimal threshold and corresponding CTI as a function of the coefficient of variation of the hyper-exponential distribution of L.

THEOREM 4.3. *An optimal Markovian policy is the threshold-based policy using threshold θ_0 such that:*

$$\theta_0 = \arg\max_{\theta} \frac{\int_l P_a(\theta + l) \, \mathrm{d}F_L(l)}{\int_l \log\left(1 + \frac{l}{\theta}\right) \, \mathrm{d}F_L(l)} . \tag{3}$$

PROOF. We denote with $f(\theta) = \int_l P_a(\theta + l) \, \mathrm{d}F_L(l)$ and $g(\theta) = \int_l \log\left(1 + \frac{l}{\theta}\right) \, \mathrm{d}F_L(l)$. Note that clearly $f(\theta)$ and $g(\theta)$ are non-negative functions, with $g(\theta) > 0 \; \forall \theta > 0$. It can be shown that:

$$\frac{\int_\theta f(\theta) \, \mathrm{d}\pi_U(\theta)}{\int_\theta g(\theta) \, \mathrm{d}\pi_U(\theta)} \leq \frac{f(\theta_0)}{g(\theta_0)} .$$

Indeed,

$$\frac{\int_\theta f(\theta) \, \mathrm{d}\pi_U(\theta)}{\int_\theta g(\theta) \, \mathrm{d}\pi_U(\theta)} = \frac{\int_\theta [f(\theta) - f(\theta_0)] \, \mathrm{d}\pi_U(\theta) + f(\theta_0)}{\int_\theta [g(\theta) - g(\theta_0)] \, \mathrm{d}\pi_U(\theta) + g(\theta_0)} = \frac{f(\theta_0)}{g(\theta_0)} \left(\frac{1 + \int_\theta \left(\frac{f(\theta)}{f(\theta_0)} - 1 \right) \, \mathrm{d}\pi_U(\theta)}{1 + \int_\theta \left(\frac{g(\theta)}{g(\theta_0)} - 1 \right) \, \mathrm{d}\pi_U(\theta)} \right) .$$

Now we get the assertion by proving that

$$\frac{1 + \int_\theta \left(\frac{f(\theta)}{f(\theta_0)} - 1 \right) \, \mathrm{d}\pi_U(\theta)}{1 + \int_\theta \left(\frac{g(\theta)}{g(\theta_0)} - 1 \right) \, \mathrm{d}\pi_U(\theta)} = \frac{1 + \alpha}{1 + \beta} \leq 1.$$

Of course the assertion is trivially true if $\beta > \alpha$, i.e., $\alpha - \beta \leq 0$. The latter expression holds since by construction: $\alpha - \beta = \int_\theta \left(\frac{f(\theta)}{f(\theta_0)} - \frac{g(\theta)}{g(\theta_0)} \right) \, \mathrm{d}\pi_U(\theta)$ and, by definition of θ_0,

$$\frac{f(\theta)}{f(\theta_0)} - \frac{g(\theta)}{g(\theta_0)} = \frac{g(\theta)}{f(\theta_0)} \left[\frac{f(\theta)}{g(\theta)} - \frac{f(\theta_0)}{g(\theta_0)} \right] \leq 0.$$

Notice that $\theta_0 < \infty$ is a consequence of our assumptions on P_a and $\{\tau_i\}_i$. □

Fig. 6 shows the impact of the distribution of L (specifically, the variance of L) on the optimal threshold and the corresponding maximum CTI. We consider for L a simple hyper-exponential

distribution of the second order, which allows us to vary the coefficient of variation (COV) while keeping the mean fixed to 1. We further assume that $P_a = 0.1e^{-u}$ and $\alpha = 0.1$, as in the example of Fig. 5. Interestingly, as the variance of L increases, the optimal value of the threshold decreases while the resulting CTI increases. This can be explained as follows: by increasing the coefficient of variation of L, we observe few (rare) larger and larger spikes of L, interleaved with many smaller and smaller spikes. While large (rare) spikes of L quickly fade away thanks to the exponential decay of user excitation, the presence of many small spikes allows the system to sample the user excitation at a lower level, thus yielding larger values of user response.

4.2 Real-time delivery

Recall that in this scenario the system has to make an instantaneous binary decision (selection) of impression opportunities arriving according to a Poisson process of rate λ.

The optimal selection policy can be formalized as a Markov decision process over continuous space. The state of the process is given by the user excitation sampled at the time instants at which a new impression opportunity arrives. At each sampling time t_n, two decisions a are possible: either the opportunity is sent to the user ($a = 1$), or it is discarded ($a = 0$). Thus, the instantaneous reward at the generic sampling time t_n is:

$$\mathcal{R}(n, a) = \begin{cases} P_a(U(t_n) + L_n) & \text{if} \quad a = 1 \\ 0 & \text{if} \quad a = 0 \end{cases} \tag{4}$$

For tractability, we approximate the above Markov decision process by discretizing the level of user excitation and defining a Markov Chain where the i-th state corresponds to excitation level U_i. So doing, we can apply known results from the theory of stochastic dynamic programming [29, Ch. 5], which allows us to characterize the optimal filtering policy. In particular, Theorem 2.4 in [29, Ch. 5] states that, if the Markov Chain, for any possible policy, includes an ergodic state, the policy that maximizes the average reward is stationary, i.e., the action taken at a given time instant deterministically depends on the current state. Note that, in our system, the state corresponding to any arbitrarily small level ϵ of user excitation is ergodic since it can be reached from any other state due to the decaying behavior of the excitation and the fact that L is assumed to have finite moments.

In general, the optimal policy can be found by solving Belman's equation [29]:

$$w + h(n) = \max_a \left[\mathcal{R}(n, a) + \sum_{j=0}^{\infty} P_{ij}(a)h(j) \right], \quad n \geq 0$$

where $h(n)$ is a bounded function, w is a constant representing the average optimal reward and $P_{ij}(a)$ is the probability to move from state i to j, given that decision a is made.

By using this approach, we found that in many cases of practical interest the optimal policy is threshold-based: impression opportunities arriving at time instants at which $U(t_n) \leq \theta^*$, where θ^* is an optimized threshold, have to be delivered to the user, whereas when $U(t_n) > \theta^*$ opportunities have to be discarded. We remark that there are cases in which the optimal policy is not threshold-based, but we found them just in pathological cases with non-continuous P_a, and obtaining a negligible gain. We will show one of these counter-examples in Appendix C.

Thus, in the following we will focus on the performance achieved by threshold-based policies. For given θ, we can compute the stationary distribution $\pi_U(u)$ induced by the corresponding threshold-based policy. Indeed, for real-time delivery, the CTI resulting from a threshold-based

policy can be related to $\pi_U(u)$ as:

$$\text{CTI} = \lambda \int_0^\theta P_a(u+l) f_L(l) \, \mathrm{d}\pi_U(u).$$

Let $f_U(u)$ be the probability density function of U (i.e., $f_U(u) = \frac{\mathrm{d}\pi_U(u)}{\mathrm{d}u}$).

THEOREM 4.4. *Function $f_U(u)$ is the only normalized solution (such that $\int f_U(u) \, \mathrm{d}u = 1$) of the integral equation:*

$$f_U(u) = \lambda \int_0^\infty e^{(\alpha-\lambda)t} \int_0^\theta f_L(ue^{\alpha t} - x) f_U(x) \, \mathrm{d}x \, \mathrm{d}t + \lambda \int_{\frac{1}{\alpha} \log(\frac{\theta}{u})}^\infty f_U(ue^{\alpha t}) e^{(\alpha-\lambda)t} \, \mathrm{d}t.$$

PROOF. To obtain the stationary distribution $\pi_U(u)$, we derive the Chapman-Kolmogorov equations associated to the Markov process (over continuous space state) $U_i = U(t_i)$:

$$U_{i+1} = (U_i + L_i \mathbf{1}_{[U_i \leq \theta]}) e^{-\alpha \tau_{i+1}}$$

where $\tau_{i+1} = t_{i+1} - t_i$ and the condition in the indicator function accounts for the fact that an impression opportunity is delivered only if $U_i \leq \theta$. Now, denoted with $f_{U_i}(u)$ the probability density function of U_i, we can derive an integral equation relating $f_{U_i}(u)$ to $f_{U_{i+1}}(u)$. To do so, let $V_i = (U_i + L_i \mathbf{1}_{[U_i \leq \theta]})$. By conditioning on the value x assumed by U_i, we have:

$$f_{V_i}(u \mid x) = f_L(u-x) \mathbf{1}_{[x \leq \theta]} + \delta(u-x) \mathbf{1}_{[x > \theta]}$$

where the first and second term on the right hand side account, respectively, for the case where an impression opportunity is delivered ($x \leq \theta$) and the case where the user excitation is above the threshold. Then, unconditioning, we get:

$$f_{V_i}(u) = \int_0^\theta f_L(u-x) f_{U_i}(x) \, \mathrm{d}x + f_{U_i}(u) \mathbf{1}_{[u > \theta]}.$$

Now we can observe that, conditionally over $\tau_{i+1} = t$, U_{i+1} and V_i are deterministically related, being $U_{i+1} = V_i e^{-\alpha t}$. Therefore, $f_{U_{i+1}}(u \mid t) = e^{\alpha t} f_{V_i}(ue^{\alpha t})$. Then

$$f_{U_{i+1}}(u \mid t) = e^{\alpha t} \int_0^\theta f_L(ue^{\alpha t} - x) f_{U_n}(x) \, \mathrm{d}x + e^{\alpha t} f_{U_i}(ue^{\alpha t}) \mathbf{1}_{[ue^{\alpha t} > \theta]}.$$

Finally, unconditioning we get:

$$f_{U_{i+1}}(u) = \lambda \int_0^\infty e^{(\alpha-\lambda)t} \int_0^\theta f_L(ue^{\alpha t} - x) f_{U_i}(x) \, \mathrm{d}x \, \mathrm{d}t + \lambda \int_{\frac{1}{\alpha} \log(\frac{\theta}{u})}^\infty f_{U_i}(ue^{\alpha t}) e^{(\alpha-\lambda)t} \, \mathrm{d}t. \tag{5}$$

Given the assumption on stationarity, we can impose that $f_{U_i}(u) = f_{U_{i+1}}(u) = f_U(u)$, thus $f_U(u)$ is obtained as the only normalized solution of the stated integral equation. □

4.3 Buffer-Driven Filtering (BDF) strategy

We now consider the buffer delivery scenario lying in between the extreme cases of arbitrary delivery and real-time delivery. As anticipated in Sec. 3, we assume that, similarly to the arbitrary delivery case, the Ad Server can send impressions to the user at arbitrary time instants (the user is supposed to be permanently exposed to ads), but within a given deadline for each ad. In other words, ads can be buffered for some time D (deterministic or random) at the Ad Server. We denote by $\rho = \lambda \mathbb{E}[D]$ the corresponding 'traffic intensity'. For such intermediate scenario, we propose a heuristic strategy, named buffer-driven filtering (BDF), based on the following idea: whenever there are at least two ads in the buffer, we employ the optimal threshold θ_0 of arbitrary delivery. When we have a single ad in the buffer, we employ the optimal real-time threshold θ^*. Moreover,

we delay the deliver of this single ad until it is close to expire, in order to maximize its acceptance probability. The detailed pseudo-code is provided in Appendix D.

4.4 Numerical evaluation

We now compare the performance of the strategies presented in the previous sections. To this end, we consider a scenario in which we fix again $\alpha = 0.1$, while L is assumed to be exponentially distributed (with mean 1). We consider either a monotonically decreasing response function, or an inverse-U response function. As we vary the arrival rate λ of the arriving (Poisson) stream of (candidate) impression opportunities, we compare the CTI achieved by:

- the stateless thinning strategy (SLT), with optimal thinning probability p;
- the CTI achievable under arbitrary delivery, hereinafter denoted by $CTI_0 = CTI(\theta_0)$, which represents an upper-bound to the system performance (Sec. 4.1);
- the threshold-based real-time strategy, with optimal threshold θ^* (Sec. 4.2);
- the buffer-driver filtering (BDF) strategy introduced above, for different values of traffic intensity ρ.
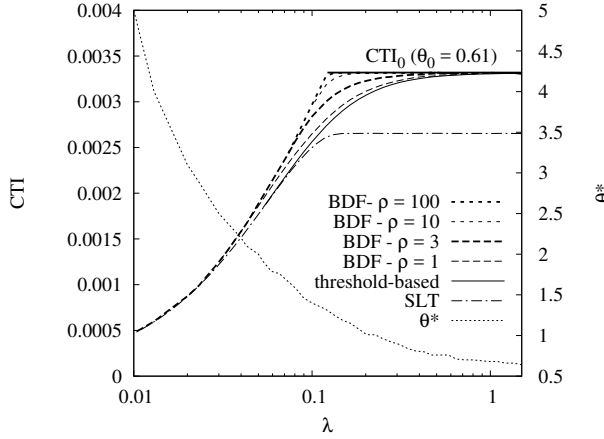

Fig. 7. CTI vs. impression rate λ under various strategies, for monotonically decreasing response function $P_a = 0.1e^{-u}$.

Fig. 7 and Fig. 8 show the CTI achieved by the considered strategies in the case of, respectively, decreasing response function $P_a = 0.1e^{-u}$ and inverse-U response function $P_a = 0.1ue^{-u}$. In both figures we also show on the right y axes the value of the optimal threshold θ^* for the real-time delivery case.

Note that, consistently with what said in Sec. 3.4, the upper bound CTI_0 provided by arbitrary delivery only makes sense when $\lambda > \lambda_{\min}$, where $\lambda_{\min}$ is the rate of impression opportunities delivered to the user by the optimal policy. Such minimum rate is given by:

$$\lambda_{\min} = \frac{\alpha}{\int_l \log\left(1 + \frac{l}{\theta_0}\right) dF_L(l)}$$

This explains why we show CTI_0 as an horizontal line starting from the point at which $\lambda = \lambda_{\min}$. Specifically, we have $CTI_0 = 0.0033$ in the case of Fig. 7, for $\theta_0 = 0.61$, whereas we get $CTI_0 = 0.005$ in the case of Fig. 8, for $\theta_0 = 1.47$.
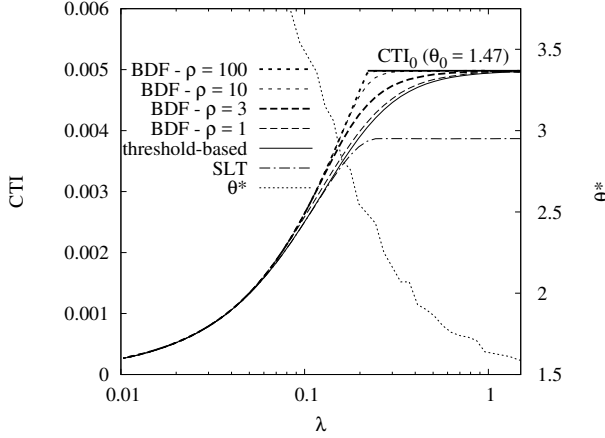
Fig. 8. CTI vs. impression rate λ under various strategies, for inverse-U response function $P_a = 0.1ue^{-u}$.

Results are qualitatively similar for both considered user response functions. All strategies tend to achieve similar performance for small λ, where filtering is not needed and the best choice is to just deliver to the user all arriving impression opportunities.

As λ increases and filtering becomes effective, some differences arise. The CTI obtained by the threshold-based policy increases with λ, till it saturates to the upper-bound. Note that the corresponding value of optimal threshold in the case of real-time delivery, θ^*, decreases with λ, approaching θ_0. The SLT strategy, instead, saturates to a lower value and much earlier than the threshold-based policy. The reason is that SLT is a simpler strategy unaware of the system state, and for large λ cannot do any better than delivering a Poisson stream of opportunities to the user (which is suboptimal).

The performance of the BDF algorithm strongly depends on the traffic intensity ρ (note that ρ equals the average buffer occupancy of an M/G/∞ queue storing the arriving opportunities). As expected, for large values of ρ (e.g., larger than 10), BDF approaches the upper bound, while for small values of ρ (e.g., smaller than 1) BDF essentially behaves like the real-time delivery strategy.

4.5 Trace-driven results

In Sec. 3.3 we already described a simple way to fit our model to the Avazu and Outbrain traces, assuming fixed $L = 1$ and obtaining an empirical user response function with $\alpha = 0.3$. We now make an additional step, comparing the actual CTI measured on the traces with the CTI resulting from the fitted model. Specifically, for both traces we consider a simple least-square exponential fitting of the user response function (the dashed lines in Figs. 3(a) and 3(b)). Using the actual arrival process of impressions appearing in the traces, we build the excitation $U(t)$ of each user (assuming that users are homogeneous), setting $\alpha = 0.3$ and $L = 1$. We can then obtain the CTI as predicted by the model, and compare it with the actual CTI of the traces. Despite the strong approximations introduced by our methodology (e.g., the least-square fitting, the assumption that users are homogeneous), we found that, for both traces, the CTI predicted by the model closely matches the actual CTI, with a relative error below 5%. This further confirms the validity of the system model introduced in Sec. 3.

We then tried to apply our filtering strategies to the arrival process of ads contained in the traces, in order to achieve possible better CTI. Unfortunately, the arrival rate of impressions in both traces is significantly smaller than $\lambda_{\min}$, leaving little room for improvements. For example, we tried the

simple SLT strategy, with a numerically optimized thinning probability p. The optimal p turns out to be equal to 1, meaning that SLT does not lead to any improvement. Due to the scarcity of impression opportunities, the arbitrary delivery strategy cannot be applied ($\lambda < \lambda_{\min}$). For the same reason, the threshold-based real-time strategy yields only marginal gains (in the order of 1%), since $U(t)$ very rarely goes above the optimal threshold θ^*.

When, instead, the BDF strategy is adopted (assuming $\rho = 100$), i.e., impressions can be delayed, and thus better spread out over time, we obtained significant improvements in CTI, namely, 7% in the case of the Avazu trace and 24% for the Outbrain trace. This can be explained by the fact that the arrival process of ads in both traces is quite bursty.

In conclusion, we can say that: i) real traces confirm the validity of our model in representing user behavior and predicting the resulting CTI; ii) for moderate values of impression rate, a careful optimization of the times at which impressions are submitted to the user (when this is possible, i.e., when ads can be delayed) can produce significant gains in terms of CTI.

5 ESTIMATION OF THE EXCITATION INFORMATION

We now move to the case in which the Ad Server does not have perfect knowledge of the user excitation at time t, $U(t)$. However, the following information is available to the advertising platform: (i) the time instants at which previous impression opportunities have been delivered to the user (i.e., the sequence $\{t_i\}_{i:t_i<t}$), and (ii) the outcomes $\{X_i\}_{i:t_i<t}$, of such opportunities (i.e., whether the user has performed a valuable action ($X_i = 1$) or not ($X_i = 0$), on each opportunity). We also assume that the system is aware of the statistics of L_i, of α and of the user-response function P_a. Notice that if a user is lowly active, i.e., rarely online and rarely clicking on ads, it will be hard to correctly estimate the user peculiar parameters (P_a, α, and L_i) and the excitation level $U(t)$. However, in this scenario, there would be little room for improvement in CTI.

From the above information, the Ad Server can derive an estimate $\hat{U}(t)$ of the exact value of user excitation $U(t)$, to be used for deciding when/whether to send the next opportunity to the user. In this case of partial excitation information, the estimation of $\hat{U}(t)$ can then be used in all the strategies presented in Sec. 4 for solving the different ads delivery scenarios.

The best estimate $\hat{U}(t)$ is, by construction, given by $\hat{U}(t) = \mathbb{E}[U(t) \mid \{t_i\}_{i:t_i<t}, \{X_i\}_{i:t_i<t}]$. Therefore, given the structure of $U(t)$, we have:

$$\hat{U}(t) = \sum_{t_i<t} \mathbb{E}[L_i \mid \{X_j\}_{t_j<t}] \exp(-\alpha(t - t_i)).$$

It follows that, in order to estimate the user excitation, we have to obtain $\mathbb{E}[L_i \mid \{X_j\}_{t_j<t}]$. The exact analysis of $\mathbb{E}[L_i \mid \{X_j\}_{t_j<t}]$ is fairly difficult. For example, if we ignored the correlation between L_i and $\{X_j\}_{j\neq i}$, we could compute $\mathbb{E}[L_i \mid X_i = 1]$ and $\mathbb{E}[L_i \mid X_i = 0]$. Then, by standard Bayesian analysis, in principle the distribution of L conditioned to the user's reaction to the offered item could be obtained as:

$$F_L(l \mid X = 1) = \frac{\int_0^l \int_0^\infty P_a(u + z) \, dF_L(z) \, d\pi_U(u)}{\int_0^\infty \int_0^\infty P_a(u + z) \, dF_L(z) \, d\pi_U(u)}.$$

and

$$F_L(l \mid X = 0) = \frac{\int_0^l \int_0^\infty (1 - P_a(u + z)) \, dF_L(z) \, d\pi_U(u)}{\int_0^\infty \int_0^\infty (1 - P_a(u + z)) \, dF_L(z) \, d\pi_U(u)}.$$

From the above expressions, it would be easy to derive $\mathbb{E}[L \mid X = 1]$ and $\mathbb{E}[L \mid X = 0]$, provided that $\pi_U(u)$ were known, which unfortunately is not.

We therefore adopt a different approach which aims at estimating directly the distribution $\pi_U^{(i)}(u)$ associated with the i-th impression opportunity sent to the user. To this end, we can write a recursive equation which relates $\pi_U^{(i)}(u)$ to $\pi_U^{(i-1)}(u)$, given X_i.

Let us define $U(t_i^+)$ as the user response at time t_i^+, i.e., right after the excitation increment that occurs upon the delivery of an impression opportunity; more formally, $U(t_i^+) = \lim_{t \downarrow t_i} U(t)$. Also, for brevity, we denote by U_i and $\pi_U^{(i)}(u)$, respectively, $U(t_i^+)$ and the distribution of $U(t_i^+)$. Using such notation, we have:

$$\mathbb{P}(U_i < u \mid X_i = 1, U_{i-1} = y, \tau_i) = \frac{\mathbb{P}(U_i < u, X_i = 1 \mid U_{i-1} = y, \tau_i)}{\mathbb{P}(X_i = 1 \mid U_{i-1} = y, \tau_i)}.$$

with:

$$\mathbb{P}(U_i < u, X_i = 1 \mid U_{i-1} = y, \tau_i) = \int_0^{(u-u_i^-)^+} P_a(u_i^- + l) \, \mathrm{d}F_L(l).$$

where $u_i^- = ye^{-\alpha\tau_i}$ and $(z)^+ = \max(0, z)$. Furthermore,

$$\mathbb{P}(X_i = 1 \mid U_{i-1} = y, \tau_i) = \int_0^\infty P_a(u_i^- + l) \, \mathrm{d}F_L(l).$$

Therefore,

$$\pi_U^{(i)}(u \mid X_i = 1, \tau_i) = \mathbb{P}(U_i < u \mid X_i = 1, \tau_i) = \frac{\mathbb{P}(U_i < u, X_i = 1 \mid \tau_i)}{\mathbb{P}(X_i = 1 \mid \tau_i)}$$

$$= \frac{\int_0^\infty \mathbb{P}(U_i < u, X_i = 1 \mid U_{i-1} = y, \tau_i) \, \mathrm{d}\pi_U^{(i-1)}(y)}{\int_0^\infty \mathbb{P}(X_i = 1 \mid U_{i-1} = y, \tau_i) \, \mathrm{d}\pi_U^{(i-1)}(y)}$$

$$= \frac{\int_0^\infty \int_{l \in [0,(u-ye^{-\alpha\tau_i})^+]} P_a(ye^{-\alpha\tau_i} + l) \, \mathrm{d}F_L(l) \, \mathrm{d}\pi_U^{(i-1)}(y)}{\int_0^\infty \int_0^\infty P_a(ye^{-\alpha\tau_i} + l) \, \mathrm{d}F_L(l) \, \mathrm{d}\pi_U^{(i-1)}(y)}. \qquad (6)$$

Similarly,

$$\pi_U^{(i)}(u \mid X_i = 0, \tau_i) = \mathbb{P}(U_i < u \mid X_i = 0, \tau_i)$$

$$= \frac{\int_0^\infty \int_0^{(u-ye^{-\alpha\tau_i})^+} (1 - P_a(ye^{-\alpha\tau_i} + l)) \, \mathrm{d}F_L(l) \, \mathrm{d}\pi_U^{(i-1)}(y)}{\int_0^\infty \int_0^\infty (1 - P_a(ye^{-\alpha\tau_i} + l)) \, \mathrm{d}F_L(l) \, \mathrm{d}\pi_U^{(i-1)}(y)}. \qquad (7)$$

The above equations can be used to construct at time t_i^+ an estimate of $\pi_U^{(i)}(u)$, given τ_i, $\pi_U^{(i-1)}(u)$ and X_i. Note that, while the information about X_i is exploited to estimate $\pi_U^{(i)}(u)$, the same information is ignored when the estimate of $\pi_U^{(i-1)}(u)$ is built. Therefore, a natural refinement of our estimate exploits the information about X_i to obtain an a-posteriori estimate of $\pi_U^{(i-1)}(u)$. To this end, we can write a "backward" equation that gives us $\mathbb{P}(U_{i-1} < y, X_i = 1 \mid U_i = u, X_{i-1}, \tau_i)$. First, observe that $\mathbb{P}(U_{i-1} < y, X_i = 1 \mid U_i = u, X_{i-1}, \tau_i) = \mathbb{P}(U_{i-1} < y, \mid U_i = u, X_{i-1}, \tau_i)$, since X_i and U_{i-1} are conditionally independent, given U_i, and $\mathbb{P}(X_i = 1 \mid U_i = u, X_{i-1}, \tau_i) = 1$. Therefore, we have:

$$\mathbb{P}(U_{i-1} < y \mid, U_i = u, X_{i-1}, \tau_i) = \frac{\mathbb{P}(U_{i-1} < y, X_{i-1} \mid U_i = u, \tau_i)}{\mathbb{P}(X_{i-1} \mid, U_i = u, \tau_i)}.$$

with:

$$\mathbb{P}(U_{i-1} < y, X_{i-1} = 1 \mid U_i = u, \tau_i) = \frac{1}{F_L(u)} \int_{(u-ye^{-\alpha\tau_i})^+}^u P_a\left((u - l)e^{\alpha\tau_i}\right) \mathrm{d}F_L(l).$$

Furthermore,

$$\mathbb{P}(X_{i-1} = 1 \mid U_i = u, \tau_i) = \frac{1}{F_L(u)} \int_0^u P_a\left((u - l)e^{\alpha \tau_i}\right) \mathrm{d}F_L(l)$$

Then we get:

$$\pi_U^{(i-1)}(y \mid X_{i-1} = 1, \tau_i) = \mathbb{P}(U_{i-1} < y \mid X_{i-1} = 1, \tau_i)$$

$$= \frac{\int_0^\infty \int_{(u-ye^{-\alpha \tau_i})^+}^u P_a\left((u - l)e^{\alpha \tau_i}\right) \mathrm{d}F_L(l) \, \mathrm{d}\pi_U^{(i)}(u)}{\int_0^\infty \int_0^u P_a\left((u - l)e^{\alpha \tau_i}\right) \mathrm{d}F_L(l) \, \mathrm{d}\pi_U^{(i)}(u)}. \quad (8)$$

and similarly:

$$\pi_U^{(i-1)}(y \mid X_{i-1} = 0, \tau_i) = \mathbb{P}(U_{i-1} < y \mid X_{i-1} = 0, \tau_i) =$$

$$\frac{\int_0^\infty \int_{(u-ye^{-\alpha \tau_i})^+}^u \left(1 - P_a\left((u - l)e^{\alpha \tau_i}\right)\right) \mathrm{d}F_L(l) \, \mathrm{d}\pi_U^{(i)}(u)}{\int_0^\infty \int_0^u \left(1 - P_a\left((u - l)e^{\alpha \tau_i}\right)\right) \mathrm{d}F_L(l) \, \mathrm{d}\pi_U^{(i)}(u)}. \quad (9)$$

We can therefore iterate between ((6) or (7)) and ((8) or (9)), until convergence to a fixed point is reached.

Note that, from $\pi_U^{(i)}(u)$, we can easily obtain the distribution of $U(t)$ for any $t \in (t_i, t_{i+1})$ since, by construction, we have: $U(t) = u_i e^{-\alpha(t - t_i)}$ and, hence,

$$\mathbb{P}(U(t) < u) = \mathbb{P}(U_i < ue^{\alpha(t - t_i)}) = \pi_U^{(i)}(ue^{\alpha(t - t_i)}).$$

In particular,

$$\mathbb{P}(U(t_{i+1}^-) < u) = \mathbb{P}(U_i < ue^{\alpha(t_{i+1}^- - t_i)}) = \pi_U^{(i)}(ue^{\alpha \tau_{i+1}}).$$

Notice that the methodology presented here could be computationally heavy. In order to reduce its time complexity, integrals could be computed a-priori by discretizing in space U and time ΔT.

5.1 Numerical results

In the previous section we described a methodology to estimate the user excitation $U(t)$, given the uncertainty on the sequence of L_i. Aiming at showing the performance of this method, hereinafter referred to as *Feedback*, we exploit (6)–(9) to compute $\mathbb{E}[U(t)]$ and use it to decide when/whether to deliver an impression opportunity to the user. We then derive the resulting CTI for both arbitrary and real-time delivery. To this end, we keep the same setting as in Sec. 4.4, i.e., $\alpha = 0.1$, L exponentially distributed with mean 1, and $P_a = 0.1e^{-u}$; also, we set $\lambda = 1$ in the real-time delivery case. We compare *Feedback* to a simpler approach, called *Fixed-L*, according to which the evolution of $U(t)$ is estimated by assuming L to be deterministic (equal to 1).

In the arbitrary delivery scenario, we apply the optimal threshold-based policy and compute the error in the estimate of the user excitation and the consequent loss in terms of CTI with respect to the case where perfect information is available (CTI_0). Fig. 9 depicts the average value of such relative loss and highlights that *Feedback* gets closer to the performance achieved in the presence of perfect information.

In the case of real-time delivery, we apply the threshold-based, real-time strategy and we report the relative gain we obtain with respect to the SLT method with optimal p. Fig. 10 depicts such a gain, along with the results achieved when perfect knowledge of U is available. Again, *Feedback* better approximates the case of perfect information and improves the SLT performance by almost 17%.

Then we assume that the Ad Server has inaccurate knowledge of the decay rate α of user excitation, namely, it underestimates α by 50%. Consequently, the Ad Server will compute an inaccurate value of the threshold, hence it will expose the user to impressions at non-optimal values of user excitation. The performance of *Fixed-L* and *Feedback* in such scenario is illustrated in Fig. 9 and 10. As expected, the performance decreases with respect to the case in which the exact α is known; however, *Feedback* can better cope with the incorrect knowledge of α since it is able to partially compensate the errors by exploiting the information from user response.
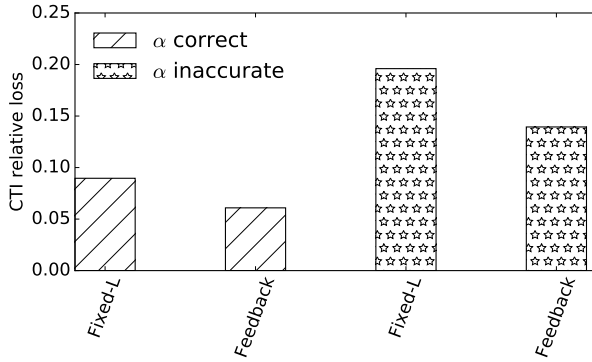

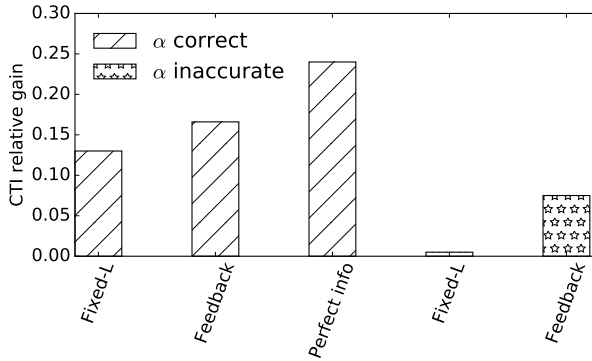Fig. 9. Relative CTI loss in arbitrary delivery scenario, under different approximations of U.

Fig. 10. Relative CTI gain in real-time delivery scenario, under different approximations of U.

6 DISCUSSION AND CONCLUSIONS

In this work we have explored the novel problem of maximizing the number of actions per time unit of a user subject to impressions. Although our analysis relies on a specific user behavioral model, traces of real advertising systems have confirmed the existence of significant correlations between the click probability and the history of impressions shown to the user, which are well captured by our approach. Indeed, by fitting the parameters of the model on the considered traces, we have been

Table 2. Strategies for different delivery scenarios. All the strategy can be applied with perfect or partial excitation information.

Ads delivery scenario	Proposed strategy	Notes	Section
Arbitrary	SLT with optimal p	Baseline	3.5, Appendix A
	Threshold-based Markovian	Optimal	4.1
Real time	Markovian, solve Belman's equation	Optimal	4.2
	Threshold-based Markovian	Often optimal	4.2
Buffer	BDF, hybrid (arbitrary/real time strategies)	Sub-optimal	4.3, Appendix D

able to accurately predict the CTI resulting from the temporal sequence of impressions appearing in the traces. More importantly, the model has allowed us to optimize the sequence of impressions itself, achieving significant gains in terms of CTI (and thus in terms of profits). In Table 2, we summarize the proposed strategies for the different delivery scenarios and refer to the Sections where they were presented. All the strategies can be applied even in case of partial excitation information, by estimating the excitation with the Feedback strategy presented in Sec. 5.

As a first step in this new research problem, our analysis has adopted several simplifying assumptions which also have important limitations, as we briefly discuss below suggesting directions for future work. First, we have assumed that candidate impressions are homogeneous (i.e., they are qualitatively identical). A natural extension of our analysis would be to consider heterogeneous impressions, for example divided into different classes having their own arrival rate and user response function. In this case the solution should control the cumulative excitation received by the user. Second, for analytical tractability we have often assumed that the arrival process of impression opportunities is Poisson. Since the activity patterns of a user might not be well described by a simple Poisson process [31], it would be desirable to extend the analysis to more realistic processes, possibly taking into account the on-off nature (sessions) of user exposure. Third, it would be interesting to perform a wider sensitivity analysis on the various parameters of the model, to better assess the impact of incomplete information about the user characteristics. At last, in the real world a user is usually exposed to several competing Ad Servers and ad campaigns, and its response to them is typically not independent, but driven by cumulative effects. A game-theoretic approach would be appropriate to analyze such competition scenario.

ACKNOWLEDGMENTS

REFERENCES

[1] Adwords help: Frequency capping (website). https://support.google.com/google-ads/answer/117579. Accessed July 2018.
[2] Avazu dataset. https://www.kaggle.com/c/avazu-ctr-prediction. Accessed July 2018.
[3] Iab's internet advertising revenue report (website). https://www.iab.com/wp-content/uploads/2019/05/Full-Year-2018-IAB-Internet-Advertising-Revenue-Report.pdf. Accessed September 2019.
[4] Outbrain dataset. https://www.kaggle.com/c/outbrain-click-prediction. Accessed July 2018.
[5] E. Altman. Game theoretic approaches for studying competition over popularity and over advertisement space in social networks. Technical report, INRIA, 2012. https://hal.inria.fr/hal-00683781/.

[6] K. Bawa. Modeling inertia and variety seeking tendencies in brand choice behavior. *Marketing Science*, 9(3):263–278, 1990.

[7] D. Berlyne. *Conflict, Arousal and Curiosity*. McGraw-Hill, 1960.

[8] J. T. Cacioppo and R. E. Petty. Effects of message repetition and position on cognitive response, recall, and persuasion. *Journal of Personality and Social Psychology*, 37(1):97–109, 1979.

[9] B. J. Calder and B. Sternthal. Television commercial wearout: An information processing view. *Journal of Marketing Research*, 17(2):173–186, May 1980.

[10] H. Cao, E. Chen, J. Yang, and H. Xiong. Enhancing recommender systems under volatile userinterest drifts. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 1257–1266, New York, NY, USA, 2009. ACM.

[11] O. Celma and P. Herrera. A new approach to evaluating novel recommendations. In *Proceedings of the 2008 ACM Conference on Recommender Systems*, RecSys '08, pages 179–186, New York, NY, USA, 2008. ACM.

[12] J. Chen and J. Stallaert. An economic analysis of online advertising using behavioral targeting. *MIS Q.*, 38(2):429–450, June 2014.

[13] P. Cremonesi, F. Garzotto, S. Negro, A. Papadopoulos, and R. Turrin. Comparative evaluation of recommender system quality. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, pages 1927–1932, New York, NY, USA, 2011. ACM.

[14] A. Das Sarma, S. Gollapudi, R. Panigrahy, and L. Zhang. Understanding cyclic trends in social choices. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, pages 593–602, New York, NY, USA, 2012. ACM.

[15] Y. Ding and X. Li. Time weight collaborative filtering. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM '05, pages 485–492, New York, NY, USA, 2005. ACM.

[16] M. Elahi, F. Ricci, and N. Rubens. A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*, 20:29 – 50, 2016.

[17] P. Farris, B. Neil, P. Phillip, and R. David. *Marketing Metrics: The Manager's Guide to Measuring Marketing Performance*. Pearson Education, 2016.

[18] D. G. Goldstein, R. P. McAfee, and S. Suri. The cost of annoying ads. *SIGecom Exch.*, 13(2):47–52, Jan. 2015.

[19] N. Gupta, A. Das, S. Pandey, and V. K. Narayanan. Factoring past exposure in display advertising targeting. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 1204–1212, New York, NY, USA, 2012. ACM.

[20] K. Kapoor, K. Subbian, J. Srivastava, and P. Schrater. Just in time recommendations: Modeling the dynamics of boredom in activity streams. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM '15, pages 233–242, New York, NY, USA, 2015. ACM.

[21] Y. Koren. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 447–456, New York, NY, USA, 2009. ACM.

[22] T. Q. Lee, Y. Park, and Y.-T. Park. A time-based approach to effective recommender systems using implicit feedback. *Expert Syst. Appl.*, 34(4):3055–3062, May 2008.

[23] W. Lu, S. Ioannidis, S. Bhagat, and L. V. Lakshmanan. Optimal recommendations under attraction, aversion, and social influence. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 811–820, 2014.

[24] X. Lu, X. Zhao, and L. Xue. Is combining contextual and behavioral targeting strategies effective in online advertising? *ACM Trans. Manage. Inf. Syst.*, 7(1):1–20, Feb. 2016.

[25] A. Maksai, F. Garcin, and B. Faltings. Predicting online performance of news recommender systems through richer evaluation metrics. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, pages 179–186, New York, NY, USA, 2015. ACM.

[26] S. M. McNee, J. Riedl, and J. A. Konstan. Being accurate is not enough: How accuracy metrics have hurt recommender systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, pages 1097–1101, New York, NY, USA, 2006. ACM.

[27] J. M. J. Murre and J. Dros. Replication and Analysis of Ebbinghaus' Forgetting Curve. *PLOS ONE*, 10(7), 07 2015.

[28] C. Rohrer and J. Boyd. The rise of intrusive online advertising and the response of user experience research at yahoo! In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, pages 1085–1086, New York, NY, USA, 2004. ACM.

[29] S. Ross. *Introduction to Stochastic Dynamic Programming*. Academic Press, 1995.

[30] J. Shanahan and D. den Poel. Determining optimal advertisement frequency capping policy via markov decision processes to maximize click through rates. In *Proceedings of NIPS Workshop: Machine Learning in Online Advertising*, pages 39–45, 2010.

[31] L. Vassio, M. Mellia, F. Figueiredo, A. P. Couto da Silva, and J. M. Almeida. Mining and modeling web trajectories from passive traces. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 4016–4021, Dec 2017.

[32] L. Vassio, H. Metwalley, and D. Giordano. The exploitation of web navigation data: Ethical issues and alternative scenarios. In F. D'Ascenzo, M. Magni, A. Lazazzara, and S. Za, editors, *Blurring the Boundaries Through Digital Innovation*, pages 119–129, Cham, 2016. Springer International Publishing.

[33] R. W. Wolff. *Stochastic modeling and the theory of queues*. Pearson College Division, 1989.

[34] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen. How much can behavioral targeting help online advertising? In *WWW '09*, pages 261–270, New York, NY, USA, 2009. ACM.

A OPTIMAL SLT UNDER POISSON STREAM

Let $M_U(\theta) = \mathbb{E}_U[e^{\theta U(t_n)}]$ the moment generating function associated to the stationary distribution of $U(t_n)$. Since $\{t_n\}_n$ is a Poisson process, we have $M_U(\theta) = \mathbb{E}_U[e^{\theta U(0)}]$, since by PASTA we can consider any arbitrary time instant (namely, 0). Now, by construction:

$$U(0) = \sum_{t_n \in (-\infty, 0)} L_n e^{\alpha t_n}$$

We also define the truncated version:

$$U_T(0) = \sum_{t_n \in [-T, 0)} L_n e^{\alpha t_n}$$

The moment generating function of $U_T(0)$, denoted by M_U^T, can be easily obtained by conditioning on the number of points of $\{t_n\}$ that fall in $[-T, 0)$. Therefore, we can exploit the following property of Poisson processes: conditionally over their number, the non-ordered points, Z_i, of a Poisson process falling over a finite interval are uniformly and independently distributed. It follows that:

$$M_U^T(\theta \mid m) = \mathbb{E}\left[e^{\theta \sum_{i=1}^m L_i e^{\alpha Z_i}}\right] = \mathbb{E}\left[\prod_{i=1}^m e^{\theta L_i e^{\alpha Z_i}}\right] = \left(\mathbb{E} e^{\theta L_1 e^{\alpha Z_1}}\right)^m. \tag{10}$$

with:

$$\mathbb{E}\left[e^{\theta L_1 e^{\alpha Z_1}}\right] = \frac{1}{T} \int_{-T}^{0} \int_l e^{\theta l e^{\alpha z}} \, \mathrm{d}F_L(l) . \, \mathrm{d}z$$

Therefore,

$$M_U^T(\theta \mid m) = \left(\frac{1}{T} \int_{-T}^{0} \int_l e^{\theta l e^{\alpha z}} \, \mathrm{d}F_L(l) \, \mathrm{d}z\right)^m.$$

and, unconditioning:

$$M_U^T(\theta) = \sum_{m=0}^{\infty} \left(\frac{1}{T} \int_{-T}^{0} \int_l e^{\theta l e^{\alpha z}} \, \mathrm{d}F_L(l) \, \mathrm{d}z\right)^m \frac{(\lambda p T)^m}{m!} e^{-\lambda p T}$$

$$= \exp\left(-\lambda p T + \lambda p \int_{-T}^{0} \int_l e^{\theta l e^{\alpha z}} \, \mathrm{d}F_L(l) \, \mathrm{d}z\right)$$

$$= \exp\left(-\lambda p \int_{-T}^{0} \int_l (1 - e^{\theta l e^{\alpha z}}) \, \mathrm{d}F_L(l) \, \mathrm{d}z\right).$$

Then for $T \to \infty$, we get:

$$M_U(\theta) = \exp\left(-\lambda p \int_{-\infty}^{0} \int_l (1 - e^{\theta l e^{\alpha z}}) \, \mathrm{d}F_L(l) \, \mathrm{d}z\right)$$

$$= \exp\left(-\lambda p \int_{0}^{\infty} \int_l (1 - e^{\theta l e^{-\alpha z}}) \, \mathrm{d}F_L(l) \, \mathrm{d}z\right). \tag{11}$$

Observe that, if we assume L to have a compact support, then for $z \to \infty$, $le^{-\alpha z} \to 0$, hence, $\exp(\theta l e^{-\alpha z}) = 1 + \theta l e^{-\alpha z} + o(e^{-\alpha z})$. In other words, the integrand function decays to 0 exponentially fast as $z \to \infty$.

Due to the complexity of (11), we aim at deriving a simpler expression for $M_U(\theta)$. To this end, for every $T > 0$, we write:

$$\int_0^T \int_l \left(1 - e^{\theta l e^{-\alpha z}}\right) dF_L(l)\, dz =$$

$$= \int_0^T \int_l \left(1 - \sum_{k=0}^{\infty} \frac{(\theta l)^k}{k!} e^{-k\alpha z}\right) dF_L(l)\, dz$$

$$= \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \int_0^T e^{-k\alpha z}\, dz \int_l l^k\, dF_L(l)$$

$$= \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \frac{1 - e^{-k\alpha T}}{k\alpha} \mathbb{E}_L[L^k]$$

Thus,

$$\int_0^{\infty} \int_l (1 - e^{l e^{-\alpha z}})\, dF_L(l)\, dz =$$

$$= \lim_{T \to \infty} \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \frac{1 - e^{-k\alpha T}}{k\alpha} \mathbb{E}_L[L^k]$$

$$= \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \frac{\mathbb{E}_L[L^k]}{k\alpha} \tag{12}$$

We therefore obtain:

$$M_U(\theta) = \exp\left(-\lambda p \sum_{k=1}^{\infty} \frac{\theta^k}{k!} \frac{\mathbb{E}_L[L^k]}{k\alpha}\right) \tag{13}$$

From (13), it is then straightforward to derive the stationary distribution $\pi_U(u)$, which will depend on p. Since the filtered arrival process is still Poisson, it follows that the optimal policy can be found by replacing the expression of $\pi_U(u)$ in the revenue and by optimizing the latter with respect to p.

B PROOF OF PROPOSITION 1

First note that, under an ergodic Markovian policy, the evolution of the user excitation is Markovian. Furthermore, it can be seen that, under any ergodic policy with stationary distribution $\pi_U(u)$, the evolution of the user excitation over continuous time forms a regenerative process. This because choosing an arbitrary $\tilde{u}$ such that $0 < \pi_U(\tilde{u}) < 1$ and defining with $\{\tilde{\sigma}_i\}_i$ the sequence of times at which $U(t) = \tilde{u}$, we can show that $\{\tilde{\sigma}_i\}_i$ forms a non defective sequence of stopping times for the continuous time process $U(t)$ (i.e., $\mathbb{E}[\tilde{\sigma}_i - \tilde{\sigma}_{i-1}] < \infty$). Indeed, let us define $\underline{S}_{\tilde{u}}$ as the set of states in correspondence of which $U(t) \leq \tilde{u}$ and $\overline{S}_{\tilde{u}}$ as the complementary set corresponding to values $U(t) > \tilde{u}$. Then we can observe that: i) by ergodicity, the sequence of times $\{\sigma_i\}_i$, at which the system enters a state in $\underline{S}_{\tilde{u}}$ from a state $\overline{S}_{\tilde{u}}$ forms a non defective sequence; ii) by construction, $\{\sigma_i\}_i \subseteq \{\tilde{\sigma}_i\}_i$, therefore $\{\tilde{\sigma}_i\}_i$ is non defective.

It follows that we can apply standard Reward-Renewal results [33], according to which the average revenue if given by

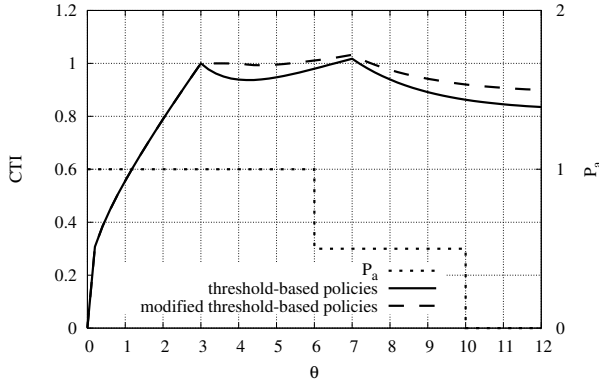$$\mathcal{R} = \frac{\mathbb{E}[P_a(U + L)]}{\mathbb{E}[\tau]} \tag{14}$$

Fig. 11.   Counterexample showing that the optimal policy for real-time delivery is not always threshold-based.

where τ is the intertime between two generic impression opportunities to which the use is exposed. In order to derive the denominator, we express the user excitation at time instant t_{i+1} (time at which the user is exposed to impression opportunity $i + 1$), as a function of the user excitation at time instant t_i (time at which the user has been exposed to opportunity i):

$$U(t_{i+1}) = (U(t_i) + L_i)e^{-\alpha \tau_{i+1}} \tag{15}$$

where $\tau_{i+1} = t_{i+1} - t_i$. Thus, we get:

$$\tau_{i+1} = \frac{1}{\alpha} \log \frac{U(t_i) + L_i}{U(t_{i+1})} \tag{16}$$

from which we obtain:

$$\begin{aligned} \mathbb{E}[\tau] &= \frac{1}{\alpha} \mathbb{E}_L \mathbb{E}_U \left[\log \left(U(t_i) + L_i \right) - \log U(t_{i+1}) \right] \\ &= \frac{1}{\alpha} \int_l \int_u \left[\log(u + l) - \log u \right] \mathrm{d}\pi_U(u) \, \mathrm{d}F_L(l) \\ &= \frac{1}{\alpha} \int_l \int_u \log \left(1 + \frac{l}{u} \right) \mathrm{d}\pi_U(u) \, \mathrm{d}F_L(l) \, . \end{aligned} \tag{17}$$

By replacing (17) into (14), we get the thesis.

C   EXAMPLE OF NON-OPTIMALITY OF THRESHOLD BASED POLICY IN REAL-TIME DELIVERY

We fix $\lambda = 2$, $L = 3$, and response function P_a shown by the dotted curve in Fig. 11. Note that the chosen P_a contains two sharp discontinuities at $U = 6$ and $U = 10$. Employing a discretization step $\Delta u = 0.1$, we found that the optimal policy filters out all impression opportunities arriving at time instants at which either i) $U(t_n) > 7$ or ii) $U(t_n) \in [3, 4.1]$, obtaining CTI ≈ 1.03. The best threshold-based policy requires $\theta^* = 7$, and achieves CTI ≈ 1.02. This is shown in Fig. 11, which reports the CTI achieved by the whole family of threshold-based policies, as we vary θ (solid line). Fig. 11 also shows the CTI achieved by the same family of threshold-based policies, modified by discarding opportunities arriving in the 'hole' $U(t_n) \in [3, 4.1]$, which contains the optimal one for $\theta^* = 7$.

The above counterexample has to be considered an academic curiosity: in all cases that we tested where P_a is continuous, like in the exponential case, the optimal policy is threshold based. Moreover, the above counterexample suggests that possible gains achievable by non-threshold-based policies are negligible (e.g., CTI $\approx$ 1.03 vs CTI $\approx$ 1.02).

D THE BDF STRATEGY

We report below the pseudo-code of the BDF strategy, assuming that the optimal threshold θ_0 (θ^*) is available for arbitrary (real-time) delivery. The algorithm maintains two state variables: the buffer length q (in number of impression opportunities), and a boolean flag *saturation* taking value 1 while operating as in arbitrary delivery, value 0 while operating as in real-time delivery. We assume that the system maintains on online estimate $\hat{U}(t)$ of the user excitation $U(t)$. The code shows the operations performed upon the occurrence of three main events: i) the arrival of a new impression opportunity; ii) $\hat{U}$ drops below θ_0; iii) an opportunity kept in the buffer expires. Note that, when the strategy decides to send an impression to the user, it chooses the one with the earliest deadline (Earliest Deadline First, EDF), so as to maximize the buffer occupancy.

Algorithm 1 BDF strategy

Require: θ_0, θ^*
  $q = 0$
  saturation = 0
  $\hat{U} \leftarrow$ *initial estimate of U*
  upon event < arrival of impression opportunity > **do**
    if $q = 1$ **then**
      if $\hat{U} < \theta_0$ **then**
        send-impression-to-user (EDF)
        $\hat{U} \leftarrow$ *update estimate of U*
        if $\hat{U} < \theta_0$ **then**
          saturation = 0
        else
          saturation = 1
      else
        saturation = 1
        $q \leftarrow 2$
    else
      enqueue-impression
      $q \leftarrow (q + 1)$
  upon event < $\hat{U}$ drops below θ_0 > **do**
    if *saturation* = 1 & $q > 0$ **then**
      send-impression-to-user (EDF)
      $q \leftarrow (q - 1)$
    if $q < 2$ **then**
      saturation = 0
  upon event < impression opportunity expires > **do**
    if $q = 1$ & $\hat{U} < \theta^*$ **then**
      send-impression-to-user
    else
      discard-impression-opportunity
    $q \leftarrow (q - 1)$
    if $q < 2$ **then**
      saturation = 0