Sustaining a Good Impression: Mechanisms for Selling Partitioned Impressions at Ad-Exchanges*

Sameer Mehta, Milind Dawande, Ganesh Janakiraman, Vijay Mookerjee
The University of Texas at Dallas, Richardson, Texas 75080

{sameer.mehta, milind, ganesh, vijaym}@utdallas.edu

In the mobile advertising ecosystem, the role of ad-exchanges to match advertisers and publishers has grown significantly over the past few years. At a mobile ad-exchange, impressions (i.e., opportunities to display ads) are sold to advertisers in real time through an auction mechanism. The traditional mechanism selects a single advertiser whose ad is displayed over the entire duration of an impression, i.e., throughout the user's visit. We argue that such a mechanism leads to an *allocative inefficiency*, as displaying only the winning ad throughout the lifetime of an impression precludes the exchange from exploiting the opportunity to obtain additional revenue from advertisers whose willingness-to-pay becomes higher *during* the lifetime of that impression. Our goal in this paper is to address this efficiency loss by offering mechanisms in which multiple ads can be displayed sequentially over the lifetime of the impression. We consider two plausible settings – one, where each auction is individually rational for the advertisers and the other, where advertisers are better off relative to the traditional mechanism over the long run – and derive an optimal (i.e., revenue-maximizing for the ad-exchange) mechanism for each setting. To efficiently compute the payment rule, the optimal mechanism for the former setting uses randomized payments. Under this mechanism, while the ad-exchange always benefits relative to the traditional mechanism, the advertisers could either gain or lose – we demonstrate both these possibilities. The optimal mechanism for the latter setting is a "mutually-beneficial" mechanism in that it guarantees a win-win for both the parties relative to the traditional mechanism, over the long run. Happily, for both the mechanisms, the allocation of ads and the payments from the advertisers are efficiently computable, thereby making them amenable to real-time bidding.

Key words: Mobile Advertising, Ad-Exchanges, Optimal Mechanisms, Mutually-Beneficial Mechanisms

---

1. Introduction

Many mobile publishers, i.e., owners of mobile applications or *apps*, earn revenue via advertisements on their apps. This revenue stream has sustained primarily due to the tremendous growth in mobile advertising over the past few years; in FY 2017, mobile advertising accounted for 56.7% ($49.9 billion) of the total advertising revenue (Interactive Advertising Bureau 2017). One of the key drivers of this growth is the emergence of multiple channels to sell digital ads. Until recently, publishers would sell most of their impressions through ad-networks[1] via long-term contracts. These contracts

* Forthcoming in **Information Systems Research**.

[1] Ad-networks (or supply-side networks) aggregate impressions from different publishers and sell them to advertisers.

are usually drawn on a revenue-sharing basis, where the ad-network shares a proportion of the revenue it earns with the publishers.

A now-popular way to buy and sell digital ads is via an *ad-exchange* – an online, automated marketplace that connects advertisers and publishers to buy and sell ads through auctions in real time[2]; examples include DoubleClick, RightMedia, AppNexus, and OpenX. Ad-exchanges are attractive to publishers as they provide liquidity in selling impressions, offer better transparency than ad-networks, and help elicit better prices from advertisers. Also, advertisers get access to a large inventory of impressions and are able to better target their audience. With benefits to both publishers and advertisers, the growth of ad-exchanges to trade digital ads has surged over recent years. Present-day ad-exchanges are closely aligned with publisher goals – higher revenue for the ad-exchange translates to higher revenue for the publishers.

The focus of our study is on display advertising on a mobile device; e.g., a smartphone or a tablet. At a mobile ad-exchange, advertisers – or equivalently, ad agencies who manage ad campaigns on behalf of advertisers – bid for impressions originating from mobile devices; DoubleClick, PubMatic, Smaato, and Nexage, are some prominent present-day mobile ad-exchanges. As will become clear soon, mobile in-app advertising can especially benefit from the kind of inefficiency we seek to eliminate, as app sessions typically last much longer than user visits on webpages – 4.2 minutes on average, as compared to just under 1 minute, according to a recent study (iAd 2014). The traditional auction for click ads that has been used in these exchanges solicits a bid from each advertiser (for the value he derives from the user clicking on that ad) in a real-time auction; the winner's ad is displayed on the impression over the entire lifetime of that impression; i.e., throughout the user's visit. We highlight an allocative inefficiency in the traditional auction via a simple but illustrative example.

1.1. Inefficiency in Traditional Allocation

Consider the traditional auction with two advertisers (1 and 2), competing to display their ads, say A and B, respectively, on a mobile impression. Suppose advertiser 2 wins the auction, as a result of which ad B is displayed on the user's app. As the app session progresses, the click-probability of ad B varies due to a variety of factors. Since the value of displaying an ad to its advertiser is tied to its click probability, the (instantaneous) value of ad B to its advertiser also varies with the passage of time. This notion of the varying value of ad B is represented by the dotted curve in Figure 1.

[2] See https://developers.google.com/ad-exchange/rtb/start for more details on real-time bidding at Google's DoubleClick Ad-Exchange.

Now, consider ad A that lost to ad B in the auction and is, therefore, not displayed on the user's app. Had advertiser 1 won the auction and ad A were displayed to the user, its (instantaneous) value would also vary with the passage of time but possibly at a different rate than that of ad B. This is represented by the solid curve in Figure 1. Notice that in time interval $[t_1, t_2]$, the (instantaneous) value of displaying ad A is higher than that of ad B; thus, during this interval, advertiser 1 would be willing to pay more for displaying ad A to the user than would advertiser 2 to display ad B. However, since advertiser 2 won the auction, ad B is displayed to the user *throughout the lifetime* of the impression! Thus, the shaded region in Figure 1 represents, from the ad-exchange's perspective, an allocative inefficiency[3] from the use of the traditional auction.

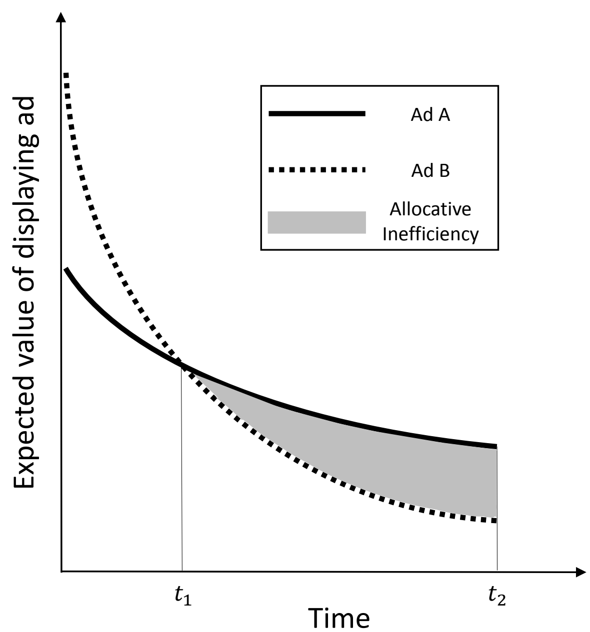


Figure 1 Allocative inefficiency in the traditional auction at ad-exchanges: An illustrative example.

This brings us to the idea of allocating (possibly) multiple ads to an impression *when it is auctioned*, with each allocated ad being scheduled for display in specific "slots" of time; of course, the ad scheduled for display at a certain time can actually be shown only if the impression lasts until that time. For instance, in the above example, the mechanism used by the ad-exchange could choose to allocate ad A to the interval $[0, t_1]$, and ad B to the interval $[t_1, t_2]$. Allocating multiple

[3] The mechanism-design literature studies several sources of inefficiencies that arise in the design of auctions – e.g., no trade as a consequence of reserve prices set by the principal or the resource not being allocated to the highest bidder when the bidders are asymmetric. However, the focus of our work is on the allocative inefficiency that arises from the principal allocating the *entire resource at once to a single bidder*.

ads to an impression potentially gives the ad-exchange an opportunity to obtain more surplus from the trade. Further, there is no technological hurdle in implementing this idea – e.g., the DoubleClick and OpenX exchanges allow publishers to dynamically reload ads on an impression[4]. Thus, a natural goal is an effective mechanism for the exchange to sell such "partitioned" impressions.

Given the discussion above, a mechanism that immediately suggests itself is one that considers each slot separately and allocates the best ad in that time slot. However, such a mechanism will yield sub-optimal revenue to the ad-exchange as it does not take into account the impact of the ad placed in the current time slot on the (potential) time slots in the future. We will analyze this mechanism and show that the revenue from this "myopic" mechanism can be significantly lower than the optimal mechanism, and it is therefore not an attractive prospect for the ad-exchange.

A related issue is the welfare implication of such a mechanism on the advertisers. While the ad-exchange will clearly benefit relative to the traditional mechanism, it is not immediately clear whether the advertisers will too. If the advertisers could be worse-off in some situations, then one could also consider alternate mechanisms under which the ad exchange ensures that the advertisers are at least as well-off as they are under the traditional mechanism.

Given that a predominant volume of impressions are being sold today via ad-exchanges, the removal of this allocative inefficiency by designing attractive mechanisms has significant implications for the advertising ecosystem. Moreover, such a mechanism should be suited for real-time bidding, i.e., the allocation of ads and the payment from the advertisers should be efficiently computable. For wider acceptability, it would also help if the mechanism is simple in its structure. We address all these goals in this study.

Before proceeding further, we briefly discuss our main contributions.

1.2. Our Contributions

We propose an approach to address the allocative inefficiency in present-day mobile ad-exchanges that arises from the sale of an impression to a single advertiser – as discussed above, the allocation rule of the traditional mechanism fails to exploit advertisers' willingness-to-pay over the lifetime of the impression. To this end, we develop a framework that addresses this efficiency loss, namely the partitioning of an impression into time slots and allocating (possibly) multiple ads sequentially to these time slots when the impression is auctioned.

We obtain optimal mechanisms for selling partitioned impressions under two plausible settings that differ in terms of the utility guarantee they offer to the advertisers: The first (Setting 1)

[4] See https://support.google.com/adxseller/answer/6286179 for more details.

considers mechanisms in which *each* auction is individually rational for the advertisers. Given that advertisers typically participate in tens of thousands of such auctions each day, the *long-term utility* of each advertiser is also a reasonable metric. Accordingly, our second setting (Setting 2) considers mechanisms in which advertisers obtain at least as much utility as in the traditional mechanism, over the long term. We obtain optimal mechanisms under each of these settings – the one for Setting 1 is referred to as OPT-IR (optimal individually-rational) and the one for Setting 2 as OPT-MB (optimal mutually-beneficial). We note that the traditional mechanism is feasible under both the settings.

The OPT-IR and the OPT-MB mechanisms are well-suited for real-time bidding. The mechanisms consist of an allocation rule (assignment of ads to time slots) and a payment rule (amounts charged to the advertisers for displaying their respective ads). An ad-exchange needs to process several activities from the time it receives an impression to the time it delivers the ad to the publisher's app. This includes packaging information about the impression (user demographics, OS platform type, publisher app type, etc.) into several categories, revealing all or part of this information to advertisers, obtaining bids from advertisers, conducting an auction, and finally displaying the winning advertiser's ad on the impression. The entire process must be completed fast enough (usually within 150 milliseconds), so that the app user does not experience any perceptible delay in the rendering of the ad. The allocation rules of both the optimal mechanisms can be efficiently obtained by solving a deterministic dynamic program of complexity $\mathcal{O}(AN)$, where A is the number of advertisers in the auction and N is the maximum number of time slots that the impression is partitioned into. Thus, the allocation rules can be incorporated into the ad-delivery process without causing a significant delay. Further, the allocation rule of the OPT-MB mechanism is first-best; i.e., generates the maximum possible social welfare through the trade.

The payment rule of the OPT-IR mechanism, however, is difficult to compute "as is" for the following reason. This mechanism needs to compute the information rent to each advertiser, which, in turn, is obtained from the allocation rule of the mechanism. For our problem, the allocation rule consists of a *sequence* of ads (i.e., an assignment of ads to time slots). With the increase in the bid of an ad, both the *sequence* of ads displayed in the time slots as well as the *subset of slots* assigned to that ad can change several times. Further, there is no characterization of *when* the sequence of ads changes. This makes the payment rule of the OPT-IR mechanism difficult to compute and thus poses an implementation challenge. As a remedy, we develop a *randomized* payment rule that is both optimal and easy to implement. Further, the payment can be implemented in either of the two most popular formats in use today – cost-per-click (CPC) or cost-per-impression (CPM). The payment rule of the OPT-MB mechanism can be easily computed. This simplification stems from the fact that the OPT-MB mechanism is not required to compute the information rent; it provides each advertiser exactly his expected utility from the traditional mechanism, which can be pre-computed. Therefore, in the OPT-MB mechanism, each advertiser pays his value obtained from the allocation less his expected utility from the traditional mechanism.

We also analyze the welfare implications of the OPT-IR and the OPT-MB mechanisms on the ad-exchange and the advertisers. Since the mechanism-design problem in each of the two settings is formulated from the perspective of the ad-exchange and the traditional mechanism is feasible for this problem, it is clear that the ad-exchange benefits from the optimal mechanism relative to the traditional mechanism. However, advertisers may be worse off under the OPT-IR mechanism relative to the traditional mechanism. We demonstrate both the possibilities – i.e., the OPT-IR mechanism benefiting (win-win) and hurting (win-lose) the advertisers – analytically. On the other hand, each advertiser's utility under the OPT-MB mechanism is, by construction, at least that under the traditional mechanism, in the long run.

Finally, we examine the performance of the OPT-IR and the OPT-MB mechanisms on an illustrative suite of instances and contrast their relative strengths. The main message from this numerical study is that both the mechanisms can effectively address the allocative inefficiency in the traditional mechanism – the ad-exchange can gain handsomely (ranging from 7% to 33% in our study) under either mechanism, and this gain further improves as the ads become more heterogeneous in terms of their click-probabilities over time. Moreover, both the mechanisms are attractive in terms of the social welfare they generate – the OPT-IR mechanism achieves a near-first-best social welfare and the OPT-MB mechanism achieves the first-best social welfare.

Organization of the paper: We review the relevant literature in Section 2. In Section 3, we first discuss preliminaries that are common to both Setting 1 and Setting 2, and then formulate the mechanism-design problem for Setting 1. In Section 4, we derive an optimal mechanism (OPT-IR) for Setting 1 and address a variety of issues related to its implementation. Section 5 analyzes a "myopic" mechanism that conducts an auction in each time slot and Section 6 compares its performance with the OPT-IR mechanism. Next, Section 7 analyzes the traditional mechanism that is currently used by mobile ad-exchanges. Section 8 compares the OPT-IR mechanism with the traditional mechanism by evaluating its impact on advertisers; here we show analytically that advertisers can either gain or lose under the OPT-IR mechanism. In Section 9, we obtain an optimal mechanism (OPT-MB) for Setting 2; this mechanism guarantees a higher revenue to the ad-exchange and a higher surplus to the advertisers, relative to the traditional mechanism. Section 10 investigates the OPT-IR and the OPT-MB mechanisms numerically and quantifies their benefit over the traditional mechanism. Section 11 concludes.

2. Literature Review

The design of advertising auctions in the digital ecosystem has been studied extensively in the literature. For example, Edelman et al. (2007), Varian (2007), Liu et al. (2010), and Abhishek and Hosanagar (2013), design mechanisms for slots/positions in sponsored-search auctions (see Chapter 28 of Nisan et al. 2007 for an excellent discussion on sponsored-search auctions). Kim et al. (2012) study the design problem of determining the optimal number of slots in these auctions. This stream of papers focuses on allocating (search) ads to spatial slots on a web page simultaneously, whereas we study a setting where an impression is partitioned into several time slots and (display) ads are shown sequentially. This differentiation has several bearings on the dynamics of ad allocation and payment of our mechanisms:

- Due to the sequential display of ads, the allocation rules of our mechanisms are obtained by solving a *dynamic program*.

- The display of an ad in a particular time slot is contingent on the user not leaving the app session earlier and not clicking on an ad displayed earlier in the sequence.

- In a sponsored-search auction, a single ad is allocated to at most one physical slot on the web page whereas in our mechanism, an ad can be allocated to multiple time slots.

Aumann et al. (2016) consider the problem of auctioning the timed consumption of a resource by multiple agents, where each agent has different valuations for different time intervals. For this multi-dimensional private-information setting, the social planner has to decide which agent should be allocated the resource, and for how long, to maximize total surplus (i.e., to maximize *efficiency*) instead of revenue. Another difference with respect to our work is that they only allow allocation of contiguous time intervals to the agents. As agents have different valuations for different time intervals, they show that the associated allocation problem is NP-complete and therefore seek computationally-efficient mechanisms. In contrast, our problem deals with *optimal* mechanisms that allocate time slots of predetermined length to advertisers with single-dimensional private information.

McAfee and Vassilvitskii (2012) describe a broad set of practical issues – efficiency, expressiveness, strategic simplicity, and neutrality towards participants – in designing exchanges. The mechanisms that we derive in our study are guided by these principles. Specifically, we develop revenue-maximizing mechanisms for the ad-exchange that are (i) easy-to-implement in a real-time environment, (ii) incentive compatible for the advertisers, (iii) simple for advertisers to communicate, in that they only need to report their valuation-per-click. Our mechanisms also take into consideration the welfare implication on both the ad-exchange and the advertisers.

Our work is also related to the literature on scheduling display ads. The paper closest to our work is Sun et al. (2017) that analyzes the following setting. A supply-side ad-network has acquired an impression. The ad-network has a set of ads (of the advertisers that the ad-network has contracted with) that it wants to possibly display on that impression. For this purpose, the ad-network wants to determine a sequence of ads to display on that impression during its lifetime. If an ad is clicked, the ad-network charges an ad-specific amount (valuation-per-click) to the corresponding advertiser. The advertisers sign a contract with the ad-network that specifies the amount that they have to pay to the ad-network when their respective ads are clicked. Therefore, in this setting, the ad-network has *full information* about the value that a particular ad-sequence can yield. The objective of the ad-network is to determine an ad-sequence that maximizes expected revenue. The authors obtain near-optimal algorithms (optimal under some special cases) for the ad-sequencing problem above.

The context of our paper is fundamentally different from that considered in Sun et al. (2017). In our setting, an *ad-exchange* is faced with the challenge of selling partitioned impressions to the advertisers. Unlike ad-networks, the ad-exchange does not know the per-click valuations of the advertisers, which naturally gives rise to *information asymmetry* between the ad-exchange and the advertisers. Consequently, the ad-exchange has to design a mechanism (an auction) to sell the partitioned impressions. The design of an optimal mechanism involves obtaining an allocation rule as well as a payment rule. By announcing these two rules of a mechanism to the advertisers, the ad-exchange offers appropriate incentives to the advertisers to participate in the mechanism and report their valuations truthfully in equilibrium. The optimal ad-scheduling problem studied in Sun et al. (2017) only focuses on obtaining the best schedule of ads to display on the impression and completely ignores modeling advertisers' incentives.

While the allocation rule of our optimal mechanism can be computed efficiently, the payment rule of that mechanism is not amenable to real-time implementation "as is". We address this challenge by developing the *randomized payment rule* that yields the same expected revenue as the payment rule of the OPT-IR mechanism. To the best of our knowledge, this idea of randomized payment is novel in the context of digital advertising literature and is appropriate for real-time implementation in both the CPM and CPC formats.

Another important difference is that the problem considered in Sun et al. (2017) is solely from the context of an ad-network, and completely ignores the welfare of the advertisers. In contrast, we consider the welfares of the supply-side (ad-exchange) and the demand-side (advertisers) players in our problem. This gives rise to interesting questions such as how a mechanism affects the two parties. As will be shown later, a mechanism which benefits the ad-exchange may prove harmful to the advertisers. This possibility drives the subsequent discussion in our paper, leading to the proposal of a mutually-beneficial mechanism, which is another important contribution of our paper aimed at improving current practice.

Hojjat et al. (2017) consider advertisers' constraints – reach and frequency of ads – for ad planning and delivering sequenced ads. A case study on Facebook ads (Adaptly 2014) experimentally demonstrates that creatively sequencing ads at a personalized level increases view-through and subscription rates. Mohan et al. (2013) focus on the challenges and feasibility of prefetching multiple ads in the existing advertising architecture. Bharadwaj et al. (2012) develop an efficient algorithm to allocate ads in guaranteed contracts. Turner et al. (2011) model the scheduling of ads in video games and develop a dynamic scheduling algorithm.

Several studies investigate other challenges that arise in display advertising such as: (i) targeting strategies (Goldfarb and Tucker 2011), including mobile targeting (Andrews et al. 2015, Chen et al. 2017), (ii) ad positioning (Agarwal et al. 2011), (iii) click behavior (Chatterjee et al. 2003), (iv) wearout of ads (Braun and Moe 2013), and (v) preference between the CPC and CPM pricing formats (Asdemir et al. 2012, Najafi-Asadolahi and Fridgeirsdottir 2014), among others. Goldstein et al. (2015) discuss the effectiveness of selling time-based display ads by analyzing the duration of the ads through an online behavioral experiment. Muthukrishnan (2009) and Korula et al. (2016) discuss some of the broader issues and research opportunities in the digital advertising ecosystem. Yuan et al. (2014) provide a comprehensive survey on real-time-bidding advertising.

We also note a related stream of literature that investigates supply- and demand-side issues that arise in the digital advertising ecosystem. The supply-side consists of publishers and ad-networks. Balseiro et al. (2014) model the tradeoff faced by publishers in selling their impressions via an ad-exchange for short-term revenue against the long-term benefits obtained by contracting with an ad-network. Roels and Fridgeirsdottir (2009) consider the problem of maximizing publisher revenue in the presence of advertising requests and web traffic. Balseiro et al. (2015) study various design decisions (e.g., reserve prices) in the presence of dynamic interactions among budget-constrained advertisers. Yang et al. (2010) study a publisher's problem of allocating ad-space between guaranteed delivery from an ad-network and non-guaranteed delivery from an ad-exchange, under multiple objectives for the publisher and the advertisers. The demand-side of the ecosystem primarily includes advertisers and demand-side platforms. Several papers focus on campaign-management issues from the demand-side; see, e.g., Aseri et al. (2017) and Balseiro et al. (2017). Zhang et al. (2014) derive optimal real-time bidding strategies for display advertising based on advertiser's budget, campaign objective, and impression details. Allouah and Besbes (2017) show that, under a wide range of market settings, multi-bidding by a demand-side platform – i.e., the platform submitting multiple bids to the ad-exchange instead of one – benefits both demand-side and supply-side players.

3. Preliminaries and Formulation of Setting 1

We begin by describing the key elements used in Setting 1 and Setting 2.

- **Impression**: An opportunity to display an ad on an app is referred to as an impression. We assume that time is divided into slots of equal length that is determined by industry practice as the minimum length of exposure for an ad; for instance, mobile ads sold through the DoubleClick ad-exchange are shown for a minimum of 30 seconds. Let N be a sufficiently large integer such that the length of an app session (i.e., the time the user stays on the app until he either leaves the app or clicks on an ad) can be reasonably assumed[5] to be at most N time slots; we index the time slots by n; $n = 1, 2, ..., N$. Slots of varying length can be easily accounted for by adjusting the click-probabilities of the ads in the different time slots.

- **Mobile Ad-Exchange**: For each impression that is generated on an app, the ad-exchange solicits valuation-per-click bids from advertisers for their respective ads and chooses a sequence of ads to display in the time slots of that impression. The objective of the ad-exchange is to maximize its expected revenue over the lifetime of the impression.

- **Advertisers**: We use the term "advertiser" to reference an entity that is interested in purchasing the impression. Advertisers compete to display their respective ads on the impression. We denote the set of advertisers by $\mathcal{A}$; let $|\mathcal{A}| = A$. We assume that each advertiser bids for the display of only one ad and, therefore, use the subscript a for both of them interchangeably. Let the private valuation-per-click for advertiser a be denoted by r_a, which is a random variable independently distributed with publicly-known c.d.f. $F_a(\cdot)$, and p.d.f. $f_a(\cdot)$ over the interval $\mathcal{B}_a = [0, \omega_a]$. Let $\boldsymbol{\mathcal{B}} = \times_{a=1}^{A} \mathcal{B}_a$ denote the Cartesian product of the valuation intervals of the advertisers, and for all a, let $\boldsymbol{\mathcal{B}}_{-a} = \times_{j \neq a} \mathcal{B}_j$. Let $\mathbf{r} = [r_1, r_2, ..., r_A]$ denote the vector of the true valuations and let $\mathbf{r}_{-a} = [r_1, ..., r_{a-1}, r_{a+1}, ... r_A]$. The bid submitted by advertiser a is denoted by b_a and $\mathbf{b} = [b_1, b_2, ..., b_A]$. Let $f(\mathbf{r}) = \prod_{a=1}^{A} f_a(r_a)$ denote the joint density at $\mathbf{r}$ and let $f_{-a}(\mathbf{r}_{-a})$ denote the joint density at $\mathbf{r}_{-a}$. Let $F(\mathbf{r})$ denote the joint distribution at $\mathbf{r}$.

Click-Probabilities: The click-probability of an ad a, when shown in time slot n of the impression, is denoted by $p_{a,n}$. In our analysis, we assume that the click probabilities of the advertisers are common knowledge. This is a widely-used assumption in the literature (see, e.g., Edelman et al.

[5] See Remark 2 for a mathematically-precise justification.

2007, Garg and Narahari 2009, and Thompson and Leyton-Brown 2013), with the argument that the bidders are likely to learn all relevant information about each other's ads through repeated interactions. This assumption is not too restrictive for our context, in the following sense: If, instead, the click-probabilities of an ad are private to its advertiser and the ad-exchange, and other advertisers only have distributional knowledge of these probabilities, then all our mechanisms remain incentive compatible for the advertisers. Thus, the assumption that click-probabilities are common knowledge is not needed for the advertisers to determine their bids; the assumption is only needed to establish the optimality of our mechanisms in their respective settings. For tractability, we assume that the click-probability of an ad in a slot depends only on the time elapsed thus far in the user's session. In general, the click probability of an ad in a particular slot might depend on several other characteristics, e.g., the ads previously shown to the user in that session and the sequence in which they were shown. All the optimization problems we encounter in our analysis can also be formulated using a general click-probability structure that allows this click-probability to depend on an arbitrary set of characteristics; however, then, the optimization problems are no longer efficiently solvable. Let the conditional probability that the user stays on the app until the end of a time slot, given that she enters that time slot and does not click on the ad, be denoted by λ.

We now formulate the mechanism-design problem for Setting 1.

3.1. Setting 1: Problem Formulation

Our analysis makes the following regularity assumption: For each $a \in \mathcal{A}$, the distribution $F_a(\cdot)$ with density $f_a(\cdot)$ is regular, i.e., the hazard rate $\left(\frac{f_a(r_a)}{1-F_a(r_a)}\right)$ is non-decreasing and bounded. This assumption is common in the mechanism-design literature (see, e.g., Krishna 2009) and guarantees that the *virtual* valuation, $\psi_a(r_a) := r_a - \frac{1-F_a(r_a)}{f_a(r_a)}$ is non-decreasing in r_a over the support of F_a. Distributions that satisfy this assumption include, among others, uniform, exponential, truncated normal, log-normal and Weibull. Let $\mathcal{A}^+ := \{a \mid \psi_a(r_a) \geqslant 0\} \subseteq \mathcal{A}$ denote the set of advertisers that have non-negative virtual valuations.

Using the Revelation Principle (Myerson 1981), we restrict our attention – without loss of generality – to incentive compatible (IC) and individually rational (IR) *direct* mechanisms, i.e., mechanisms in which (i) advertisers reporting their per-click valuations truthfully to the ad-exchange is a Bayesian Nash Equilibrium (BNE) and (ii) advertisers obtain a non-negative expected payoff from participating in the mechanism and are therefore willing to do so.

A direct mechanism $\boldsymbol{\mu}$ consists of a pair of functions $(\mathbf{\Pi}^{\boldsymbol{\mu}}, \mathbf{M}^{\boldsymbol{\mu}})$, where $\mathbf{\Pi}^{\boldsymbol{\mu}} : \boldsymbol{\mathcal{B}} \rightarrow \{\mathcal{A} \cup \phi\}^N$, is the *allocation rule* that specifies the ad sequence and $\mathbf{M}^{\boldsymbol{\mu}} : \boldsymbol{\mathcal{B}} \rightarrow \mathbb{R}^A$ is the *payment rule* that specifies the expected payment by the advertisers to the ad-exchange. Specifically,

- $\mathbf{\Pi}^{\boldsymbol{\mu}}(\mathbf{b}) = \left\{\pi_n^{\boldsymbol{\mu}}(\mathbf{b}) : 1 \leqslant n \leqslant N, \mathbf{b} \in \boldsymbol{\mathcal{B}}\right\}$, where $\pi_n^{\boldsymbol{\mu}}(\mathbf{b}) \in \{\mathcal{A} \cup \phi\}$ denotes the ad in time slot n for bid vector $\mathbf{b}$. We allow for a null allocation in each time slot; thus, if no ad is allocated to slot n, then $\pi_n^{\boldsymbol{\mu}}(\mathbf{b}) = \phi$ (see Remark 1 in Section 4 for an alternate formulation of the allocation function).

- $\mathbf{M}^{\boldsymbol{\mu}}(\mathbf{b}) = \left\{M_a^{\boldsymbol{\mu}}(\mathbf{b}) : a \in \mathcal{A}, \mathbf{b} \in \boldsymbol{\mathcal{B}}\right\}$, where $M_a^{\boldsymbol{\mu}}(\mathbf{b})$ denotes the expected payment made by advertiser a to the ad-exchange under the bid vector $\mathbf{b}$.

The objective of the ad-exchange is to find a mechanism $\boldsymbol{\mu}$ that maximizes its expected revenue over the N time slots of an impression. Table 1 summarizes our notation.

Notation	Parameter Description
$\mathcal{A}$	The set of all ads; $\lvert\mathcal{A}\rvert = A$.
N	The number of time slots that the impression is partitioned into.
r_a	True (private) valuation-per-click for advertiser a; $a \in \mathcal{A}$.
$\mathbf{r} = [r_1, r_2, ..., r_A]$	The valuation-per-click vector.
$f_a(\cdot), F_a(\cdot)$	The p.d.f. and c.d.f., respectively, of the valuation-per-click of advertiser a.
$f(\mathbf{r}) = \prod_{a=1}^{A} f_i(r_a)$	The joint density of vector $\mathbf{r}$.
$f_{-a}(\mathbf{r}_{-a}) = \prod_{i=1}^{a-1} f_i(r_i) \prod_{j=a+1}^{A} f_j(r_j)$	The joint density of vector $\mathbf{r}_{-a}$.
b_a	The valuation-per-click (bid) reported by advertiser a.
$\psi_a(r_a) = r_a - \frac{1-F_a(r_a)}{f_a(r_a)}$	The virtual valuation of advertiser a.
$\mathcal{A}^+$	The set of advertisers with non-negative virtual valuations.
$p_{a,n}$	Click-probability of ad a when displayed in time slot n.
λ	Conditional probability of the user staying on the app at the end of a time slot, given that she enters that time slot and does not click on the ad.

Table 1 Our Main Notation.

As is common in the mechanism-design literature, we assume that the advertisers have a quasi-linear utility function. The expected payoff to each advertiser from participating in an auction is equal to his expected value from the slots he wins minus the payment he makes to the ad-exchange. Let $\Theta_a^{\boldsymbol{\mu}}(\mathbf{b})$ denote the likelihood of a click on ad a over all the time slots, under the mechanism $\boldsymbol{\mu}$, when the advertisers bid $\mathbf{b}$. Then,

$$\Theta_a^{\boldsymbol{\mu}}(\mathbf{b}) = \sum_{n=1}^{N} p_{a,n} \mathbb{1}\left\{\pi_n^{\boldsymbol{\mu}}(\mathbf{b}) = a\right\} \lambda^{n-1} \prod_{t=1}^{n-1} (1 - p_{\pi_t^{\boldsymbol{\mu}}(\mathbf{b}),t}) \ \forall a \in \mathcal{A}. \tag{1}$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function of its argument. For the sequence $\mathbf{\Pi}^{\boldsymbol{\mu}}(\mathbf{b})$ and time slot n, the term $\lambda^{n-1} \prod_{t=1}^{n-1} (1 - p_{\pi_t^{\boldsymbol{\mu}}(\mathbf{b}),t})$ denotes the probability that in the previous $n-1$ slots, neither the user exited the app (λ^{n-1}) nor did she click on any ad $\left(\prod_{t=1}^{n-1} (1 - p_{\pi_t^{\boldsymbol{\mu}}(\mathbf{b}),t})\right)$.

The term $p_{a,n}\mathbb{1}\{\pi_n^{\boldsymbol{\mu}}(\mathbf{b})=a\}$ denotes the click-probability of ad a in time slot n. Thus, $\sum_{n=1}^{N} p_{a,n}\mathbb{1}\{\pi_n^{\boldsymbol{\mu}}(\mathbf{b})=a\}\lambda^{n-1}\prod_{t=1}^{n-1}(1-p_{\pi_t^{\boldsymbol{\mu}}(\mathbf{b}),t})$ denotes the probability that ad a is clicked when the sequence $\boldsymbol{\Pi}^{\boldsymbol{\mu}}(\mathbf{b})$ is chosen. Let $\theta_a^{\boldsymbol{\mu}}(b_a)$ denote the expected click-probability of ad a across all the slots when advertiser a bids b_a and all the other advertisers bid their true valuations. Thus, $\theta_a^{\boldsymbol{\mu}}(b_a)=\mathbb{E}_{\mathbf{r}_{-a}}\left[\Theta_a^{\boldsymbol{\mu}}(b_a,\mathbf{r}_{-a})\right]\ \forall a\in\mathcal{A}$.

Let $m_a^{\boldsymbol{\mu}}(b_a)$ denote the expected payment by advertiser a to the ad-exchange when he bids b_a and all the other advertisers report their true valuations. Thus, $m_a^{\boldsymbol{\mu}}(b_a)=\mathbb{E}_{\mathbf{r}_{-a}}\left[M_a^{\boldsymbol{\mu}}(b_a,\mathbf{r}_{-a})\right]$. Then, the net expected utility that advertiser a obtains from bidding b_a when all the other advertisers bid their true valuations is equal to $r_a\theta_a^{\boldsymbol{\mu}}(b_a)-m_a^{\boldsymbol{\mu}}(b_a)$. The IC and the IR constraints can now be stated as follows:

$$r_a\theta_a^{\boldsymbol{\mu}}(r_a)-m_a^{\boldsymbol{\mu}}(r_a)\geqslant r_a\theta_a^{\boldsymbol{\mu}}(b_a)-m_a^{\boldsymbol{\mu}}(b_a)\ \forall a\in\mathcal{A},\ \forall r_a,b_a\in\mathcal{B}_a, \tag{IC}$$

$$r_a\theta_a^{\boldsymbol{\mu}}(r_a)-m_a^{\boldsymbol{\mu}}(r_a)\geqslant 0\ \forall a\in\mathcal{A},\ \forall r_a,b_a\in\mathcal{B}_a. \tag{IR}$$

Let $U_a^{\boldsymbol{\mu}}(r_a)=r_a\theta_a^{\boldsymbol{\mu}}(r_a)-m_a^{\boldsymbol{\mu}}(r_a)$ denote the expected utility to advertiser a under mechanism $\boldsymbol{\mu}$, in equilibrium. The (IC) constraints state that it is optimal for the advertisers to reveal their private valuation-per-click truthfully, given that all the other advertisers do so. In other words, truth-telling is a BNE. The (IR) constraints state that the expected payoff of each advertiser in a BNE is non-negative.

3.2. Setting 1: The Optimal Mechanism-Design Problem

The ad-exchange solicits valuation-per-click bids from the advertisers, based on which it determines: (i) the optimal sequence (i.e., the allocation of ads to time slots) through the allocation function $\boldsymbol{\Pi}^{\boldsymbol{\mu}}$ and (ii) the expected payment from each advertiser through the payment function $\mathbf{M}^{\boldsymbol{\mu}}$. The optimization problem for the ad-exchange is:

$$\max_{\boldsymbol{\mu}}\ \left\{\sum_{a=1}^{A}\mathbb{E}_{r_a}\left[m_a^{\boldsymbol{\mu}}(r_a)\right]\right\},\quad \text{s.t. (IC), (IR)}. \tag{P$^{\text{IR}}$}$$

That is, the ad-exchange maximizes the sum of the expected payments it obtains by allocating the ads to the impression, subject to the (IC) and (IR) constraints of the advertisers.

In Section 4, we derive several solutions (i.e., optimal mechanisms) to problem (P^{IR}) that differ in their implementation. Before proceeding further, we find it convenient to discuss the related problem of identifying *efficient mechanisms* – this will be helpful in sections 4.1 and 9 in the discussions related to our optimal mechanisms.

3.3. Vickrey-Clarke-Groves (VCG) Mechanism

Consider the problem of designing an IC and IR mechanism that maximizes the *social welfare*, defined as the sum of the utilities to all the advertisers and the ad-exchange. Since payments from the advertisers to the ad-exchange are "internal transfers", the social welfare for any valuation-per-click vector $\mathbf{r}$ under an arbitrary IC mechanism $\boldsymbol{\mu}$ is

$$\mathrm{SW}^{\boldsymbol{\mu}}(\mathbf{r}) := \sum_{a=1}^{A} r_a \Theta_a^{\boldsymbol{\mu}}(\mathbf{r}). \tag{2}$$

Define $\mathrm{SW}_{-i}^{\boldsymbol{\mu}}(\mathbf{r}) := \sum_{a\neq i} r_a \Theta_a^{\boldsymbol{\mu}}(\mathbf{r})$. Using the expression of $\Theta_a^{\boldsymbol{\mu}}(\mathbf{r})$ from (1), the social welfare under mechanism $\boldsymbol{\mu}$ can be written as

$$\mathrm{SW}^{\boldsymbol{\mu}}(\mathbf{r}) = \sum_{a=1}^{A} r_a \cdot \sum_{n=1}^{N} p_{a,n} \mathbb{1}\left\{\pi_n^{\boldsymbol{\mu}}(\mathbf{r}) = a\right\} \lambda^{n-1} \prod_{t=1}^{n-1} (1 - p_{\pi_t^{\boldsymbol{\mu}}(\mathbf{r}),t}). \tag{3}$$

An *efficient*, IC and IR mechanism is one that maximizes (3) subject to (IC) and (IR). It is well-known that the class of VCG mechanisms is efficient (Krishna and Perry 1998). Specific to our setting, the VCG mechanism is described by the allocation rule $\mathbf{\Pi}^{\mathrm{VCG}}$ and the payment rule $\mathbf{M}^{\mathrm{VCG}}$ defined below:

$$\begin{aligned}
\mathbf{\Pi}^{\mathrm{VCG}}(\mathbf{r}) &= \arg\max_{\boldsymbol{\mu}} \ \left\{\mathrm{SW}^{\boldsymbol{\mu}}(\mathbf{r})\right\} \\
&= \arg\max_{\boldsymbol{\mu}} \ \left\{\sum_{a=1}^{A} r_a \cdot \sum_{n=1}^{N} p_{a,n} \mathbb{1}\left\{\pi_n^{\boldsymbol{\mu}}(\mathbf{r}) = a\right\} \lambda^{n-1} \prod_{t=1}^{n-1} (1 - p_{\pi_t^{\boldsymbol{\mu}}(\mathbf{r}),t})\right\}. \qquad (\mathrm{P}^{\mathrm{EFF}})
\end{aligned}$$

$$M_a^{\mathrm{VCG}}(\mathbf{r}) = \mathrm{SW}^{\mathrm{VCG}}(0, \mathbf{r}_{-a}) - \mathrm{SW}_{-a}^{\mathrm{VCG}}(\mathbf{r}) \ \forall a \in \mathcal{A}. \tag{4}$$

We note that problem ($\mathrm{P}^{\mathrm{EFF}}$) is identical to a problem recently studied by Sun et al. (2017) in the context of an ad-network optimizing the sequence of ads (that the network has already purchased) displayed during the lifetime of an impression. This problem can be solved using backward, dynamic programming recursion – the allocation rule, $\mathbf{\Pi}^{\mathrm{VCG}}(\mathbf{r})$, is:

$$\pi_n^{\mathrm{VCG}}(\mathbf{r}) = \arg\max_{a\in\mathcal{A}} \left\{r_a p_{a,n} + \lambda(1 - p_{a,n}) R(n+1; \mathbf{r})\right\}; \ 1 \leqslant n \leqslant N, \text{ where} \tag{5}$$

$$R(n; \mathbf{r}) = \max_{a\in\mathcal{A}} \left\{r_a p_{a,n} + \lambda(1 - p_{a,n}) R(n+1; \mathbf{r})\right\}; \ 1 \leqslant n \leqslant N, \text{ and } R(N+1; \mathbf{r}) = 0.$$

Next, we use this result and derive an optimal mechanism problem (P^{IR}).

4. Setting 1: An Optimal Mechanism

We begin by specifying an optimal mechanism for problem (P^{IR}) in Theorem 1. Then, Sections 4.1 and 4.2 discuss the challenges associated with implementing this mechanism – we address these by obtaining an optimal mechanism that uses randomized payments. We also show how these payments can be adapted to the CPC and the CPM formats. We end this section with brief remarks on an alternate allocation rule and infinite-horizon mechanisms.

Consider the following mechanism ($\mathbf{\Pi}^{\text{OPT-IR}}, \mathbf{M}^{\text{OPT-IR}}$):

$$\mathbf{\Pi}^{\text{OPT-IR}}(\mathbf{r}) = \arg\max_{\boldsymbol{\mu}} \left\{ \sum_{a=1}^{A} \psi_a(r_a)\Theta_a^{\boldsymbol{\mu}}(\mathbf{r}) \right\}, \tag{6}$$

$$M_a^{\text{OPT-IR}}(\mathbf{r}) = r_a \Theta_a^{\text{OPT-IR}}(\mathbf{r}) - \int_0^{r_a} \Theta_a^{\text{OPT-IR}}(t_a, \mathbf{r}_{-a})\, dt_a \;\; \forall a \in \mathcal{A}. \tag{7}$$

THEOREM 1. *The mechanism* ($\mathbf{\Pi}^{\text{OPT-IR}}, \mathbf{M}^{\text{OPT-IR}}$) *is an optimal solution to problem* (P^{IR}).

Proof: To show that ($\mathbf{\Pi}^{\text{OPT-IR}}, \mathbf{M}^{\text{OPT-IR}}$) is an optimal solution to problem (P^{IR}), it is sufficient to show that for every $a \in \mathcal{A}$, $\Theta_a^{\text{OPT-IR}}(\cdot, \mathbf{r}_{-a})$ is a non-decreasing function of its argument. A detailed proof of this claim is provided in Appendix A. Thus, we only need to establish that this sufficient condition holds. To this end, fix an advertiser $a \in \mathcal{A}$. Consider any two values of r_a, say $r_a^1, r_a^2 \in \mathcal{B}_a$, such that $r_a^1 \leqslant r_a^2$. Since $\psi_a(\cdot)$ is an increasing function of its argument, we have $\psi_a(r_a^1) \leqslant \psi_a(r_a^2)$. From (6), we have

$$\psi_a(r_a^2)\Theta_a^{\text{OPT-IR}}(r_a^2, \mathbf{r}_{-a}) + \sum_{j \neq a} \psi_j(r_j)\Theta_j^{\text{OPT-IR}}(r_a^2, \mathbf{r}_{-a}) \geqslant \psi_a(r_a^2)\Theta_a^{\text{OPT-IR}}(r_a^1, \mathbf{r}_{-a}) + \sum_{j \neq a} \psi_j(r_j)\Theta_j^{\text{OPT-IR}}(r_a^1, \mathbf{r}_{-a}),$$

$$\psi_a(r_a^1)\Theta_a^{\text{OPT-IR}}(r_a^1, \mathbf{r}_{-a}) + \sum_{j \neq a} \psi_j(r_j)\Theta_j^{\text{OPT-IR}}(r_a^1, \mathbf{r}_{-a}) \geqslant \psi_a(r_a^1)\Theta_a^{\text{OPT-IR}}(r_a^2, \mathbf{r}_{-a}) + \sum_{j \neq a} \psi_j(r_j)\Theta_j^{\text{OPT-IR}}(r_a^2, \mathbf{r}_{-a}).$$

Adding the two inequalities above, we get

$$\left(\psi_a(r_a^2) - \psi_a(r_a^1)\right)\left(\Theta_a^{\text{OPT-IR}}(r_a^2, \mathbf{r}_{-a}) - \Theta_a^{\text{OPT-IR}}(r_a^1, \mathbf{r}_{-a})\right) \geqslant 0.$$

Since $\psi_a(r_a^1) \leqslant \psi_a(r_a^2)$, we get $\Theta_a^{\text{OPT-IR}}(r_a^1, \mathbf{r}_{-a}) \leqslant \Theta_a^{\text{OPT-IR}}(r_a^2, \mathbf{r}_{-a})$. Therefore, $\Theta_a^{\text{OPT-IR}}(\cdot, \mathbf{r}_{-a})$ is a non-decreasing function of its argument for all a. ■

While Theorem 1 presents an optimal mechanism, our application context also requires us to implement the mechanism in real time. We now turn our attention to implementation issues.

4.1. Implementing the Optimal Allocation

Notice that the optimization problem which appears in the definition of $\mathbf{\Pi}^{\text{OPT-IR}}(\mathbf{r})$ is identical to problem (P^{EFF}) with the change that r_a is replaced by $\psi_a(r_a)$ for all $a \in \mathcal{A}$. Consequently, using the VCG-allocation specified in (5), $\mathbf{\Pi}^{\text{OPT-IR}}(\mathbf{r})$ can be computed as follows:

$$\pi_n^{\text{OPT-IR}}(\mathbf{r}) = \arg\max_{a \in \mathcal{A}^+} \left\{ \psi_a(r_a) p_{a,n} + \lambda(1 - p_{a,n}) \hat{R}(n+1; \mathbf{r}) \right\}; \; 1 \leqslant n \leqslant N, \text{ where} \tag{8}$$

$$\hat{R}(n; \mathbf{r}) := \max_{a \in \mathcal{A}^+} \left\{ \psi_a(r_a) p_{a,n} + \lambda(1 - p_{a,n}) \hat{R}(n+1; \mathbf{r}) \right\}; \; 1 \leqslant n \leqslant N, \; \hat{R}(N+1; \mathbf{r}) := 0. \tag{9}$$

The dynamic program (8) can be efficiently solved (specifically, in time $\mathcal{O}(AN)$), thus making the optimal mechanism amenable to real-time bidding. Note that, unlike sponsored-search auctions, it is possible for an ad to appear in multiple (time) slots in the sequence decided by the mechanism.

While the optimal allocation rule $\mathbf{\Pi}^{\text{OPT-IR}}(\mathbf{r})$ can be easily computed, the optimal payment rule $\mathbf{M}^{\text{OPT-IR}}(\mathbf{r})$ is difficult to evaluate – we explain and address this challenge below, and also discuss CPC and CPM implementations.

4.2. Implementing the Optimal Payment

Recall from (1) that $\Theta_a^{\text{OPT-IR}}(\mathbf{r})$ denotes the click probability of ad a across all slots when all advertisers bid truthfully, under the optimal mechanism. Therefore, the first term in the optimal payment rule ($r_a \Theta_a^{\text{OPT-IR}}(\mathbf{r})$) can be easily computed by using the optimal ad sequence obtained from (8) in the expression for $\Theta_a^{\text{OPT-IR}}(\mathbf{r})$ in (1). However, the second term in the payment rule $\left(\int_0^{r_a} \Theta_a^{\text{OPT-IR}}(t_a, \mathbf{r}_{-a})\, dt_a\right)$ involves computing $\Theta_a^{\text{OPT-IR}}(t_a, \mathbf{r}_{-a})$ for all possible bids t_a (ranging from 0 to r_a) of a. We now discuss the difficulty in computing this term.

For advertiser $a \in \mathcal{A}$, as the bid t_a changes from 0 to r_a, both the *sequence* of ads displayed in the N slots as well as the *subset of slots* assigned to advertiser a can change several times. Further, there is no characterization of *when* the sequence of ads changes. Thus, the only way to (approximately) evaluate this integral is to discretize the range $[0, r_a]$ using a sufficiently small step size, say $\Delta > 0$, and re-compute the optimal sequence r_a/Δ times, each time using the dynamic program (8). Clearly, this is computationally expensive since the step size Δ needs to be sufficiently small to ensure a reasonably-close approximation. Further, such a calculation is needed for *each* advertiser $a \in \mathcal{A}$. We now use a simple example to illustrate the change in the sequence of ads as the bid t_a of advertiser a changes from 0 to r_a.

Consider two ads X and Y and three time slots ($N = 3$). The valuation-per-click of the two ads along with their click-probabilities in the three time slots are as shown in Table 2 below. Let $\lambda = 0.9$.

	Valuation-per-click	Click-probability in time slot 1	Click-probability in time slot 2	Click-probability in time slot 3
Ad X	0.70	0.05	0.0169	0.0057
Ad Y	0.60	0.01	0.0092	0.0085

Table 2 Parameter values for ads X and Y.

To compute the payment for advertiser X, we need to evaluate $\left(\int_0^{0.70} \Theta_X^{\text{OPT-IR}}(t_X, 0.60)\, dt_X\right)$. Table 3 shows the bid values for which the optimal sequence of ads changes (and, therefore, the value of the integrand $\Theta_X^{\text{OPT-IR}}(t_X, 0.60)$ also changes).

Bid of ad X (t_X)	Ad in time slot 1	Ad in time slot 2	Ad in time slot 3
0.00	Y	Y	Y
0.53	X	Y	Y
0.56	X	X	Y
0.65	X	X	X

Table 3 Optimal sequence of ads for different bids of ad X.

To address this challenge, we introduce a novel idea of a *randomized payment rule* that yields the same expected revenue as the payment rule of the OPT-IR mechanism.

Real-Time Implementation via Randomization

Typically, advertisers bid on tens of thousands of impressions per day and thus interact with an ad-exchange on a repeated basis (Mansour et al. 2012). This motivates us to develop a randomized payment rule that can be efficiently implemented. For advertiser a, notice that

$$\begin{aligned}\int_0^{r_a} \Theta_a^{\text{OPT-IR}}(t_a, \mathbf{r}_{-a})\, dt_a &= r_a \int_0^{r_a} \Theta_a^{\text{OPT-IR}}(t_a, \mathbf{r}_{-a}) \frac{1}{r_a} dt_a \\ &= r_a \mathbb{E}_{u_a}\left[\Theta_a^{\text{OPT-IR}}(u_a, \mathbf{r}_{-a})\right], \text{ where}^6 \quad u_a \sim U(0, r_a).\end{aligned}$$

Let

$$M_a^{\text{RAND}}(\mathbf{r}) = r_a \Theta_a^{\text{OPT-IR}}(\mathbf{r}) - r_a \Theta_a^{\text{OPT-IR}}(u_a, \mathbf{r}_{-a}), \text{ where } \; u_a \sim U(0, r_a) \; \forall a \in \mathcal{A}.$$

Then, $\mathbb{E}_{u_a}\left[M_a^{\text{RAND}}(\mathbf{r})\right] = M_a^{\text{OPT-IR}}(\mathbf{r})$. This and the fact that the mechanism $(\mathbf{\Pi}^{\text{OPT-IR}}, \mathbf{M}^{\text{OPT-IR}})$ is IC implies that the mechanism $(\mathbf{\Pi}^{\text{OPT-IR}}, \mathbf{M}^{\text{RAND}})$ is IC too. Further, since the optimal allocation function

[6] The Uniform distribution is used here for ease of exposition. Any other distribution with support $[0, r_a]$ can also be used to construct such a randomized payment rule.

is monotone and $u_a \leqslant r_a$, we have $r_a \Theta_a^{\text{OPT-IR}}(\mathbf{r}) \geqslant r_a \Theta_a^{\text{OPT-IR}}(u_a, \mathbf{r}_{-a})$, which implies that the mechanism $(\mathbf{\Pi}^{\text{OPT-IR}}, \mathbf{M}^{\text{RAND}})$ is IR. Thus, the mechanism $(\mathbf{\Pi}^{\text{OPT-IR}}, \mathbf{M}^{\text{RAND}})$ is also optimal.

We now discuss the implemention of the optimal payment rule $\mathbf{M}^{\text{RAND}}$ in a manner that is consistent with the two dominant payment paradigms in digital advertising – CPM and CPC. For a traditional digital-advertising auction, where a single impression is sold to a single advertiser, the CPM model requires the winning advertiser to pay for the impression regardless of whether it is clicked on or not. The CPC model, on the other hand, requires the winning advertiser to pay for the impression only if it is clicked on. The CPM model accounts for 35% of the ad revenue in display ads while CPC accounts for 64% (Interactive Advertising Bureau 2017).

A Cost-Per-Click (CPC) Implementation

The randomized payment rule $\mathbf{M}^{\text{RAND}}$ can be implemented in a CPC-like manner by requiring an advertiser to pay only if the user clicks on that advertiser's ad. Specifically, if the user clicks on ad a, then advertiser a is required to pay $\frac{M_a^{\text{RAND}}(\mathbf{r})}{\Theta_a^{\text{OPT-IR}}(\mathbf{r})}$. This guarantees that the expected payment from any advertiser a (where the expectation is taken with respect to the randomness in whether or not the user clicks on the ad) is exactly $M_a^{\text{RAND}}(\mathbf{r})$. Thus, the allocation rule $\mathbf{\Pi}^{\text{OPT-IR}}$ along with this CPC payment rule also forms an optimal mechanism that is implementable in real time.

A Cost-Per-Impression (CPM) Implementation

A CPM-style implementation of the randomized payment scheme is to require an advertiser to pay for the display of his ad, regardless of whether or not the user clicks on the ad; the advertiser is not charged if his ad is not displayed. Here, it is important to highlight the difference with respect to CPM payments in the traditional auction. In the latter, only one advertiser wins the entire impression; thus, the winning advertiser is guaranteed that his ad will be displayed for the entire duration of the impression. In contrast, in our context, multiple advertisers might be required to pay the ad-exchange, and, with the exception of the advertiser whose ad is displayed in the first slot, the other advertisers receive no guarantee that their ads will be shown to the user – this is because the impression might end before an advertiser's turn (i.e., slot) to have his ad displayed arrives.

This motivates the following slot-by-slot CPM-style implementation of the randomized payment rule $\mathbf{M}^{\text{RAND}}$. Let $\{a \to n\}$ denote the event that ad a is displayed in slot n. Consider the payment scheme in which, if slot n materializes (that is, the impression survives until slot n), then advertiser a is required to pay $\frac{M_a^{\text{RAND}}(\mathbf{r}) \cdot \mathbb{1}\{a \to n\}}{\sum_t \text{Prob}\{a \to t\}}$ for that slot, where $\text{Prob}\{a \to t\}$ is the probability that ad a is displayed in slot t; this probability can be easily obtained from (1). Thus, the expected payment from any advertiser a over all slots (where the expectation is taken with respect to the randomness in whether or not the slot(s) assigned to this advertiser materializes) is exactly $M_a^{\text{RAND}}(\mathbf{r})$. Thus, the allocation rule $\mathbf{\Pi}^{\text{OPT-IR}}(\mathbf{r})$ along with this CPM payment rule forms an optimal mechanism.

REMARK 1. **(Alternate Allocation Rule)**: The allocation rule we formulated in Section 3.1 can be alternatively viewed as a mapping from the bid space to a probability simplex over the set of all possible sequences. Mathematically, $\mathbf{\Pi}^{\text{ALT}} : \boldsymbol{\mathcal{B}} \to [0,1]^S$, where $S = (A+1)^N$ is the number of possible ad sequences (including null allocations). It can be shown using standard arguments that this alternate formulation yields the same optimal solution as that in Theorem 1. ■

REMARK 2. **(Infinite-Horizon Mechanisms)**: While formulating our mechanism-design problem in Section 3.1, we assumed that the number of time slots, N, (that denotes the time the user stays on the app until he either leaves the app or clicks on an ad) is a sufficiently large integer. Technically, if the user neither leaves the app nor clicks on an ad, then our context requires us to consider infinite-horizon mechanisms. To show that it is sufficient to restrict attention to finite-horizon mechanisms, we establish that the loss in revenue to the ad-exchange by restricting attention to finite-horizon mechanisms instead of infinite-horizon mechanisms can be made arbitrarily small. A proof of this claim is provided in Appendix B. ■

5. A Sequential Mechanism

As discussed in Section 1, the allocative inefficiency in the traditional mechanism results from the difference in the way the expected valuations of ads change with time during an app session – ads that are more attractive for display in the initial time slots might become less attractive in the later time slots. A natural mechanism that comes to mind for addressing this allocative inefficiency is one that conducts an auction in each time slot, i.e., a mechanism that treats an impression with N potential time slots as N potential impressions and sequentially auctions each time slot independently to the advertisers. We will refer to such a mechanism as a SEQ mechanism. For a given slot $n, n \in \{1, 2, \ldots, N\}$, the optimal SEQ mechanism in that slot is defined by its allocation rule π_n^{SEQ} (ad served in time slot n) and payment rule $M_{a,n}^{\text{SEQ}}$ (amount paid by advertiser a to the ad exchange in time slot n). We further discuss this mechanism below.

Consider time slot 1. The expected value of the partitioned impression in time slot 1 to advertiser a, $a \in \mathcal{A}$, is $r_a p_{a,1}$, where $p_{a,1}$ denotes the click-probability of ad a in time slot 1. Recall that the click-probabilities of the ads in this time slot are known to the ad-exchange and only the valuation-per-click of each advertiser is private to the respective advertiser. Therefore, instead of soliciting bids for the advertisers' expected valuations $r_a p_{a,1}$, $a \in \mathcal{A}$, it suffices for the ad-exchange to solicit their valuations-per-click, $r_a, a \in \mathcal{A}$. From the theory of the classical single-unit auction (Myerson 1981), we know that there exists an optimal mechanism which is incentive-compatible and individually-rational and, further, this mechanism will serve ad $\arg\max_{a \in \mathcal{A}^+} \left\{\psi_a(r_a) p_{a,1}\right\}$ in time slot 1, where $\psi_a(r_a)$ denotes the virtual bid of advertiser a. The winning advertiser in this slot will be charged the minimum amount necessary to outbid all the other advertisers.

The session progresses to time slot 2 only if the app user does not click on the ad shown in time slot 1 and decides not to leave the app session. Conditional on the existence of time slot 2, the optimal SEQ mechanism then selects the best ad for this slot, which is ad $\arg\max_{a \in \mathcal{A}^+} \left\{\psi_a(r_a) p_{a,2}\right\}$. The winning advertiser in slot 2 pays the minimum amount necessary to outbid all the other advertisers. Repeating this argument for the N time slots, the (optimal) SEQ mechanism is defined as follows:

$$\pi_n^{\text{SEQ}}(\mathbf{r}) = \arg\max_{a \in \mathcal{A}^+} \left\{\psi_a(r_a) p_{a,n}\right\}; n = 1, 2, \ldots, N, \text{ and} \tag{10}$$

$$M_{a,n}^{\text{SEQ}}(\mathbf{r}) = r_a p_{a,n} \mathbb{1}\left\{\pi_n^{\text{SEQ}}(\mathbf{r}) = a\right\} - \int_0^{r_a} p_{a,n} \mathbb{1}\left\{\pi_n^{\text{SEQ}}(t_a, \mathbf{r}_{-a}) = a\right\} dt_a. \tag{11}$$

The SEQ mechanism defined by (10) and (11) is derived using the same steps we used to derive the OPT-IR mechanism in Appendix A. Notice that when $N = 1$, the OPT-IR mechanism and the SEQ mechanism solve the same single-slot revenue-maximization problem of allocating the ad that maximizes the expected revenue. It is then straightforward to see that the allocation rule (10) and the payment rule (11) of the SEQ mechanism in a particular time slot n, $n \in 1, 2, \ldots, N$, can be obtained, respectively, from the allocation rule (6) and the payment rule (7) of the OPT-IR mechanism.

While the SEQ mechanism conducts multiple auctions, one for each time slot, and selects an ad in each time slot, each of these auctions is structurally similar to the traditional mechanism (used by present-day ad-exchanges) that selects an ad for the entire session. Thus, the SEQ mechanism does not require any changes to the current technology infrastructure of the ad-exchange and is, consequently, easy to implement. Furthermore, since the allocation and payment rules of the SEQ mechanism are also structurally similar to those of the traditional mechanism, it would also be easy for advertisers to accommodate these rules into their bidding strategy. However, the expected revenue of the ad-exchange under the SEQ mechanism can be significantly lower than that under the optimal (i.e., OPT-IR) mechanism. The basic reason behind this drawback is as follows: While this mechanism myopically selects the "best" ad for display in the current time slot, it disregards the impact (on the ad exchange's expected revenue) from the ads that could be shown in future time slots. Nevertheless, the SEQ mechanism is attractive to practitioners due to its simplicity. Thus, it would be useful to delve deeper into the comparison of the performance of the SEQ mechanism with that of the OPT-IR mechanism. We do this next.

6. SEQ Mechanism vs. OPT-IR Mechanism

We begin our discussion by illustratating the inferior performance of the SEQ mechanism via a simple example consisting of two ads and two time slots. Throughout this section, for expositional simplicity, we consider the special case where the distribution of the valuation-per-click of the ads is a point distribution; thus, the valuation-per-click of each ad is known to the ad-exchange.

Example A: Let $\mathcal{A} = \{X, Y\}$ be the set of advertisers and $N = 2$ (the number of time slots). Let r_X and r_Y denote the valuation-per-click of advertisers X and Y, respectively. Let $r_X = r_Y\epsilon + \gamma$, where $\epsilon, \gamma > 0$, $\epsilon^2 \approx 0$, and $\gamma \approx 0$. The click-probabilities of the two ads in the two time slots are as follows: $p_{X,1} = 1, p_{X,2} = 0$ and $p_{Y,1} = p_{Y,2} = \epsilon$. Finally, let $\lambda = 1$. In time slot 1, $r_X \cdot 1 = r_Y\epsilon + \gamma > r_Y\epsilon$. Therefore, the SEQ mechanism selects ad X for display in that slot. Once ad X is displayed in slot 1, the user clicks on that ad (since $p_{X,1} = 1$) and the app session ends. Thus, the revenue to the ad-exchange under the SEQ mechanism is simply $r_X \cdot 1 = r_Y\epsilon + \gamma \approx r_Y\epsilon$. Using (6), it is straightforward to see that the OPT-IR mechanism selects ad Y for display in both the time slots. Thus, the revenue to the ad-exchange under the OPT-IR mechanism is $r_Y\epsilon + (1-\epsilon)r_Y\epsilon \approx 2r_Y\epsilon$. Thus, the ratio of the ad-exchange's revenue under the OPT-IR mechanism to that under the SEQ mechanism can be made arbitrarily close to 2. ■

In Example A, we observe that the valuation-per-click of ad Y is higher than that of ad X whereas the click-probability of ad Y in the first time slot is lower than that of ad X in the first time slot. This observation suggests that the ordering of the valuation-per-click of advertisers and the click-probability of their ads might be associated with the inferior performance of the SEQ mechanism. Before we examine this line of inquiry, it is convenient, for expositional simplicity, to define the following notion of "negative correlation": When we say that the valuation-per-click of the advertisers and the click-probability of their ads in the first time slot are "negatively correlated", we mean that the advertisers and their ads can be indexed such that:

$$r_1 \geqslant r_2 \geqslant \ldots \geqslant r_A$$
$$p_{1,1} \leqslant p_{2,1} \leqslant \ldots \leqslant p_{A,1}.$$

We begin by providing an illustrative example in which we observe that the notion of negative correlation by itself is not sufficient to guarantee the sub-optimality of the SEQ mechanism. That is, it is possible for the SEQ mechanism to be optimal when the valuation-per-click of the advertisers and the click-probability of their ads in the first time slot are negatively correlated.

6.1. Illustrative Example B

Example B: Let $\mathcal{A} = \{X, Y\}$ be the set of advertisers and $N = 2$. Let r_X (resp., r_Y) denote the valuation-per-click of advertiser X (resp, Y). Let $\lambda = 1$. The valuation-per-click and the click probabilities of each ad are shown in Table 4. Note that $r_X > r_Y$ and $p_{X,1} < p_{Y,1}$. Thus, the valuation-per-click of the advertisers and the click-probability of their ads in the first time slot are negatively correlated.

Ad	Valuation-per-click	Click probabilities	
		1	2
X	0.551	0.354	0.185
Y	0.212	0.708	0.636

Table 4 Parameter values for Example B.

For the above choice of parameters, the SEQ mechanism as well as the optimal (i.e., OPT-IR) mechanism selects ad X in time slot 1 and ad Y in time slot 2. Consequently, both the mechanisms yield a revenue of \$0.282 to the ad-exchange. ■

Next, we obtain conditions that *guarantee* the sub-optimality of the SEQ mechanism.

6.2. Sub-Optimality of the SEQ Mechanism

Let $\mathcal{A} = \{X, Y\}$ denote the set of advertisers. Let the click-probability of each ad decay at a constant rate δ with the passage of time, i.e., $p_{a,n+1} = \delta p_{a,n}$ for all $a \in \mathcal{A}$ and $n \geqslant 1$. Thus, under this constant decay-rate structure, the click-probability of an ad in any time slot is determined by its click-probability in the first slot and the decay rate. Let $\lambda = 1$. Let $\Delta(\mathbf{r}) := \frac{r_X p_{X,1} - r_Y p_{Y,1}}{p_{X,1} - p_{Y,1}}$ and define a "value function" $W(n; \mathbf{r})$ recursively as follows:

$$W(n; \mathbf{r}) := r_X \cdot p_{X,n} + (1 - p_{X,n}) \cdot W(n+1; \mathbf{r}); \;\; 1 \leqslant n \leqslant N, \;\; W(N+1; \mathbf{r}) = 0.$$

That is, $W(n; \mathbf{r})$ denotes the revenue obtained from period n onwards by displaying ad X in each time slot. Then, we have

THEOREM 2. *If $r_Y > r_X$, $p_{X,1} > p_{Y,1}$, $r_X p_{X,1} > r_Y p_{Y,1}$, and $W(K; \mathbf{r}) > \Delta(\mathbf{r}) \geqslant W(K+1; \mathbf{r})$ for some $K \in \{2, 3, \ldots, N-1\}$, then the revenue to the ad-exchange under the* OPT-IR *mechanism is strictly greater than that under the* SEQ *mechanism.*

A proof of Theorem 2 is provided in Appendix C. This result states that along with the negative correlation between the valuation-per-click of the advertisers and the click-probability of their ads in the first time slot, certain additional conditions on the problem parameters are required to guarantee the sub-optimality of the SEQ mechanism. Broadly speaking, under these conditions, the SEQ mechanism selects a single ad for display in all the time slots, whereas the OPT-IR mechanism selects two ads, and displays the first ad for a certain number of initial time slots and the second ad for the remaining time slots.

We now supplement Theorem 2 by providing a numerical example to demonstrate that the conditions in that result can indeed be achieved by the problem parameters. Consider the values of the parameters in Table 5. For these parameters, we have $\Delta(\mathbf{r}) = 0.075$. The values of $W(\cdot;\mathbf{r})$ are shown in Table 6. Thus, we have $W(3;\mathbf{r}) > \Delta(\mathbf{r}) > W(4;\mathbf{r})$ and, consequently, $K = 3$ in the statement of Theorem 2. The SEQ mechanism shows ad X in each time slot and yields a revenue of \$0.260 to the ad-exchange. On the other hand, the OPT-IR mechanism shows ad Y in the first two time slots and ad X in each of the remaining eight slots, and yields a revenue of \$0.298 to the ad-exchange. ■

N	r_X	r_Y	$p_{X,1}$	$p_{Y,1}$	δ
10	0.3	0.6	0.7	0.3	0.5

Table 5 **Parameter values for the illustrative example.**

$W(1;\mathbf{r})$	$W(2;\mathbf{r})$	$W(3;\mathbf{r})$	$W(4;\mathbf{r})$	$\ldots$	$W(10;\mathbf{r})$
0.260	0.165	0.093	0.049	$\ldots$	0.001

Table 6 **Value of $W(\cdot;\mathbf{r})$ for different time slots.**

The conditions in Theorem 2 that guarantee the sub-optimality of the SEQ mechanism are only sufficient and not necessary. We now demonstrate this via an illustrative example in which, although the valuation-per-click of two advertisers and the click-probability of their ads in the first time slot are *not* negatively correlated, the SEQ mechanism is sub-optimal.

6.3. Illustrative Example C

Example C: Consider two ads X and Y and ten time slots ($N = 10$). The click-probabilities of the ads follow a constant decay-rate structure; that is, $p_{a,n+1} = \delta_a p_{a,n}$, $a \in \{X, Y\}$ and $n \geqslant 1$. The valuation-per-click, the click-probability in the first time slot, and the decay rate of each ad are shown in Table 7. Let $\lambda = 1$. Note that, we have $r_X > r_Y$ and $p_{X,1} > p_{Y,1}$.

For the above choice of parameter values, the OPT-IR mechanism allocates ad X in the first four time slots and ad Y in the remaining six slots, and yields a revenue of \$0.884. On the other hand,

Ad	Valuation-per-click	Click probability in slot 1	Decay rate
X	0.976	0.488	0.545
Y	0.724	0.362	0.991

Table 7 Parameter values for Example C.

the SEQ mechanism allocates ad X in the first time slot and ad Y in the remaining nine slots and yields a revenue of \$0.839, which is about 5% below the optimal revenue. ■

While negative correlation, along with certain additional conditions, guarantee the sub-optimality of the SEQ mechanism, a natural question arises: When is the SEQ mechanism optimal? The following result establishes a sufficient condition.

6.4. Optimality of the SEQ Mechanism

THEOREM 3. *If* $r_1 \geqslant r_2 \geqslant \ldots \geqslant r_A$ *and* $p_{1,n} \geqslant p_{2,n} \geqslant \ldots p_{A,n}$ $\forall n \in \{1, 2, \ldots, N\}$*, then the* SEQ *mechanism is an optimal mechanism (i.e., an optimal solution to problem* P^{IR}*).*

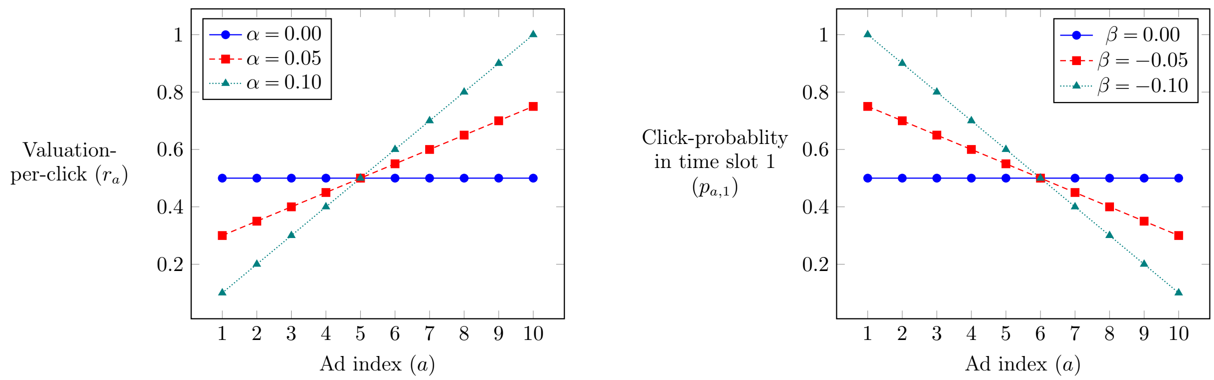
Theorem 3 states that if ads with a higher valuation-per-click have a higher click-probability in *each* time slot, then the SEQ mechanism is optimal. A proof is provided in Appendix D.

Given these developments, it should be clear that a complete characterization of the conditions under which the SEQ mechanism is optimal (or sub-optimal) is difficult to obtain. In light of this, we now conduct a focused numerical experiment to assess the impact of the *extent of negative correlation* on the performance of the SEQ mechanism relative to that of the OPT-IR mechanism. The main message here is that as the extent of negative correlation between the valuation-per-click of the advertisers and the click-probability of their ads in the first time slot decreases, the impact of the myopia of the SEQ mechanism decreases and its performance moves closer to optimality.

6.5. Numerical Analysis

Our setting is as follows: Let $\mathcal{A} = \{1, 2, \ldots, 10\}$ denote the set of advertisers and let N, the number of time slots, be equal to 10. For $a \in \mathcal{A}$, the valuation r_a of ad a is given by $r_a = 0.5 - \alpha(5 - a)$, where the parameter $\alpha \in [0, 0.1]$ is the relative difference between the valuation-per-clicks of two consecutively-indexed advertisers (see Figure 2a). Notice that a higher-indexed ad has a higher valuation-per-click than a lower-indexed ad. Let the click-probability of ad $a \in \mathcal{A}$ decay at an ad-specific rate δ_a with the passage of time, i.e., $p_{a,n+1} = \delta_a p_{a,n}$ for all $a \in \mathcal{A}$ and $n \geqslant 1$. Let the click-probability of ad $a, a \in \mathcal{A}$, in time slot 1, $p_{a,1}$, be $p_{a,1} = 0.5 - \beta(6 - a)$, where the parameter $\beta \in [-0.1, 0]$ is the relative difference between the click-probabilities of two consecutively-indexed ads in the first time slot (see

Figure 2b). Thus, lower-indexed ads have a higher click-probability in the first time slot as compared to higher-indexed ads. Let $\lambda = 1$. The decay rate δ_a of ad a is chosen randomly from a uniform distribution $U(0,1)$. For this choice of parameters, we compute the ratio of the expected revenue to the ad-exchange under the SEQ mechanism to the expected revenue to the ad-exchange under the OPT-IR mechanism, using the sample average over 1,000,000 instances.



(a) Valuation-per-click (b) Click-probability in time slot 1

Figure 2 **Illustrative figures to show (a) the relative ordering of the valuation-per-click of the ads for different values of the parameter α and (b) the relative ordering of the click-probability of the ads in time slot 1 for different values of the parameter β.**

Table 8 shows the performance of the SEQ mechanism relative to the OPT-IR mechanism for different values of the parameters α and β. When $\alpha = \beta = 0$, all the ads have the same valuation-per-click and click-probabilities in the first time slot. In this case, it is easy to verify that the SEQ mechanism is optimal and, hence, the ratio of the expected revenue of the SEQ mechanism to that of the OPT-IR mechanism is 1. As α increases and β decreases, the lower-indexed ads have increasingly lower valuation-per-click and increasingly higher click-probability in the first time slot as compared to the higher-indexed ads. That is, the extent of negative correlation among the ads increases and consequently, the performance of the SEQ mechanism deteriorates.

We now move our attention to current practice.

7. The Traditional (BASE) Mechanism

As mentioned earlier, the major mobile ad-exchanges of today use the following mechanism to sell an impression: valuation-per-click bids are solicited from the advertisers in a real-time auction and the winner gets the impression; i.e., the winner's ad is served on the impression throughout its lifetime. We will refer to this traditional mechanism as the BASE mechanism.

α	β	$\frac{\text{REVENUE(SEQ)}}{\text{REVENUE(OPT-IR)}}$
0	0	100.0%
0.01	-0.01	94.1%
0.02	-0.02	89.0%
0.03	-0.03	84.9%
0.04	-0.04	81.6%
0.05	-0.05	79.1%
0.06	-0.06	77.4%
0.07	-0.07	76.3%
0.08	-0.08	75.7%
0.09	-0.09	75.6%
0.1	-0.1	75.5%

Table 8 Impact of the parameters α and β on the performance of the SEQ mechanism relative to the OPT-IR mechanism: As the magnitude of the parameters α and β increases, the negative correlation between the valuation-per-click of the advertisers and the click-probability of the ads in the first time slot increases. This increases the extent of negative correlation among the ads and, in turn, the performance of the SEQ mechanism deteriorates.

The problem of obtaining an optimal (i.e., revenue-maximizing) BASE mechanism for the ad-exchange can be viewed as a constrained version of the problem we formulated in Section 3.1, with the additional constraint that the same ad be displayed in all the time slots. Thus, the optimization problem can be formulated as follows:

$$\max_{\boldsymbol{\mu}} \left\{ \sum_{a=1}^{A} \mathbb{E}_{r_a} \left[m_a^{\boldsymbol{\mu}}(r_a) \right] \right\} \tag{P$^{\text{BASE}}$}$$
$$\text{s.t. (IC), (IR),}$$
$$\pi_1^{\boldsymbol{\mu}}(\mathbf{r}) = \pi_2^{\boldsymbol{\mu}}(\mathbf{r}) = \ldots = \pi_N^{\boldsymbol{\mu}}(\mathbf{r}). \tag{12}$$

Using steps similar to those in Section 4 to solve problem (P$^{\text{IR}}$), the solution to problem (P$^{\text{BASE}}$) is as follows:

$$\boldsymbol{\Pi}^{\text{BASE}}(\mathbf{r}) = \arg\max_{\boldsymbol{\mu}} \left\{ \sum_{a=1}^{A} \psi_a(r_a) \Theta_a^{\boldsymbol{\mu}}(\mathbf{r}) \right\} \text{ s.t. (12),}$$
$$M_a^{\text{BASE}}(\mathbf{r}) = r_a \Theta_a^{\text{BASE}}(\mathbf{r}) - \int_0^{r_a} \Theta_a^{\text{BASE}}(t_a, \mathbf{r}_{-a})\, dt_a \;\; \forall a \in \mathcal{A}.$$

This mechanism can be simplified, using steps presented in Section E of the appendix, to yield:

$$\pi_n^{\text{BASE}}(\mathbf{r}) = \arg\max_{a \in \mathcal{A}^+} \left\{ \psi_a(r_a) \sum_{n=1}^{N} p_{a,n} \lambda^{n-1} \prod_{t=1}^{n-1} (1 - p_{a,t}) \right\}; n = 1, 2, ..., N.$$
$$M_a^{\text{BASE}}(\mathbf{r}) = \Theta_a^{\text{BASE}}(\mathbf{r}) \cdot y_a^{\text{BASE}}(\mathbf{r}_{-a}) \;\; \forall a \in \mathcal{A},$$

where $y_a^{\text{BASE}}(\mathbf{r}_{-a}) = \inf\left\{t_a : \psi_a(t_a) \geqslant 0 \text{ and } \forall\ j \neq a, \psi_a(t_a)\Theta_a^{\text{BASE}}(t_a, \mathbf{r}_{-a}) \geqslant \psi_j(r_j)\Theta_j^{\text{BASE}}(t_a, \mathbf{r}_{-a})\right\}$. In words, $y_a^{\text{BASE}}(\mathbf{r}_{-a})$ denotes the smallest bid that advertiser a has to make to win against the bid vector $\mathbf{r}_{-a}$. Thus, we have

THEOREM 4. *The mechanism* $(\mathbf{\Pi}^{\text{BASE}}, \mathbf{M}^{\text{BASE}})$ *is an optimal solution to problem* (P^{BASE}).

The optimal BASE mechanism allocates the impression to the highest bidder contingent on his virtual bid being non-negative, and charges the winner the minimum amount necessary to outbid all the other advertisers.

Our next task is to compare the OPT-IR mechanism with the BASE mechanism.

8. OPT-IR Mechanism vs. BASE Mechanism: Impact on Stakeholders

Since the BASE mechanism is a feasible solution to the mechanism design problem formulated in Section 3.1, it follows that the ad-exchange always obtains a higher expected revenue from the OPT-IR mechanism than that from the BASE mechanism. However, the surplus of an advertiser can be either lower or higher under the OPT-IR mechanism as compared to that under the BASE mechanism. We define a *win-win scenario* (signifying that both the exchange and the advertisers benefit) as one where each of the participating advertisers obtains a (weakly) higher expected utility than that under the BASE mechanism. Similarly, we define a *win-lose scenario* as one where at least one advertiser obtains a (strictly) lower expected utility than that under the BASE mechanism. We now demonstrate these two possibilities analytically.

8.1. Illustration of a Win-Win Scenario

Let $\mathcal{A} = \{X, Y\}$ be the set of advertisers and $N = 2$ (the number of time slots). Let v_X and v_Y denote the virtual bids of advertisers X and Y, respectively. Without loss of generality, we assume that the valuation-per-click bids are independently drawn from $U(0.5, 1)$. Thus, $v_X = 2r_X - 1 \sim U(0,1)$ and $v_Y = 2r_Y - 1 \sim U(0,1)$, where r_X and r_Y denote, respectively, the valuation-per-click bids of X and Y. Notice that the virtual bids are always non-negative; thus, the impression is always assigned one ad in the base mechanism and at least one ad in the optimal mechanism. Recall that $p_{a,n}$ denotes the click-probability of ad a in time slot n. Let $p_{X,1} = 2p$, $p_{X,2} = 0$, and $p_{Y,1} = p_{Y,2} = p$, where p is a parameter satisfying $0 < p \leqslant 0.5$ and $p^2 \approx 0$. Thus, ad X represents an *impulse* ad − the likelihood of a click on ad X diminishes quickly; and ad Y represents a *steady* ad − the likelihood of click on ad Y is the same in both the time slots. For expositional simplicity, assume that the conditional probability, λ, that the user stays in a particular time slot, given that she enters that time slot, is equal to 1.

For the setting described above, we evaluate the expected utility of advertisers X and Y, the expected revenue to the ad-exchange, and the expected social welfare, under the BASE and the OPT-IR mechanisms. These values are succinctly summarized in Table 9 below; for a derivation of these results, we refer the reader to Section F of the Appendix. The important observation here is that both the advertisers are better off in the OPT-IR mechanism, relative to the BASE mechanism.

	BASE	OPT-IR	Gain in OPT-IR (w.r.t. BASE)
Expected Utility of X	$\frac{p}{6}$	$\frac{7p}{24}$	+75%
Expected Utility of Y	$\frac{p}{6}$	$\frac{7p}{24}$	+75%
Expected Revenue of Ad-Exchange	$\frac{4p}{3}$	$\frac{19p}{12}$	+18.75%
Social Welfare	$\frac{5p}{3}$	$\frac{13p}{6}$	+30%

Table 9 Summary of the win-win scenario.

8.2. Illustration of a Win-Lose Scenario

Again, let $\mathcal{A} = \{X, Y\}$ and $N = 2$. The valuation-per-click bid of X, r_X, is drawn from $U(0.75, 1)$ whereas the valuation-per-click bid of Y, r_Y, is deterministically equal to 1.5. Consequently, we have $v_X = 2r_X - 1 \sim U(0.5, 1)$ and $v_Y = r_Y = 1.5$. Notice that, in this scenario too, the virtual bids of both the advertisers are always non-negative. The click probabilities of the ads in the two slots are as follows: $p_{X,1} = p_{X,2} = 1$ and $p_{Y,1} = p_{Y,2} = q$, where q is a parameter such that $0 < q \leqslant \frac{3-\sqrt{6}}{3}$. Finally, $\lambda \approx 1$. Similar to the win-win scenario, Table 10 compares the BASE and the OPT-IR mechanisms; the derivations are in Section G of the Appendix. Here, advertiser X is worse off in the OPT-IR mechanism.

	BASE	OPT-IR	Gain in OPT-IR (w.r.t. BASE)
Expected Utility of X	$\frac{1}{8}$	$\frac{(1-q)}{8}$	$-100q\%$
Expected Utility of Y	0	0	0%
Expected Revenue of Ad-Exchange	$\frac{3}{4}$	$\frac{3(1+q)}{4}$	$+100q\%$
Social Welfare	$\frac{7}{8}$	$\frac{7+5q}{8}$	$+71.43q\%$

Table 10 Summary of win-lose scenario

In summary, we have identified two scenarios, one in which shifting to the OPT-IR mechanism results in a win-win for the ad-exchange and the advertisers, relative to the BASE mechanism, and the other in which one advertiser is worse-off. The existence of these scenarios motivates us to consider mechanisms in which the benefit to both parties is *guaranteed*, i.e., the ad-exchange and the advertisers obtain, respectively, at least as much revenue and utility as in the BASE mechanism (in expectation). We refer to such mechanisms as *mutually-beneficial mechanisms* and now examine the problem of obtaining an optimal mechanism in this class.

9. Setting 2: An Optimal Mutually-Beneficial Mechanism

By definition, the BASE mechanism is mutually beneficial. Since advertisers typically bid on thousands of impressions that arrive at an ad-exchange, it is reasonable to consider mechanisms in which advertisers obtain, *over the long run*, at least their respective utilities in the BASE mechanism (which we refer to as the BASE-utility). The corresponding optimal mechanism-design problem is as follows:

$$\max_{\boldsymbol{\mu}} \left\{ \sum_{a=1}^{A} \mathbb{E}_{r_a} \left[m_a^{\boldsymbol{\mu}}(r_a) \right] \right\} \tag{P$^{\text{MB}}$}$$
$$\text{s.t. (IC)},$$
$$\mathbb{E}_{r_a} \left[U_a^{\boldsymbol{\mu}}(r_a) \right] \geqslant \mathbb{E}_{r_a} \left[U_a^{\text{BASE}}(r_a) \right] \ \ \forall a \in \mathcal{A}. \tag{13}$$

The constraints in (13) ensure that, in the long run, each advertiser obtains at least his expected utility in the BASE mechanism. Note that, problem (P$^{\text{MB}}$) relaxes the (IR) constraints in problem (P$^{\text{IR}}$) but, instead, imposes the BASE-utility constraints (13). Thus, problem (P$^{\text{MB}}$) is *not* a relaxation of problem (P$^{\text{IR}}$). Consequently, the ad-exchange's revenue under a mutually-beneficial mechanism could be higher or lower than that under the OPT-IR mechanism. Before proceeding further, we remark on our formulation of problem (P$^{\text{MB}}$).

REMARK 3. (**Individual Rationality for Individual Auctions**): The mechanism-design problem (P$^{\text{MB}}$) that we formulated above guarantees at least the BASE-utility in the long run. One could also conceive of a more-constrained setting where advertisers obtain at least the BASE-utility in the long run *and* each individual auction satisfies the individual-rationality constraints (i.e. the constraints (IR) for each auction, viz., $U_a^{\boldsymbol{\mu}}(r_a) \geqslant 0 \ \forall a \in \mathcal{A}$) for the advertisers. The mechanism-design problem for this setting can be formulated as follows:

$$\max_{\boldsymbol{\mu}} \left\{ \sum_{a=1}^{A} \mathbb{E}_{r_a} \left[m_a^{\boldsymbol{\mu}}(r_a) \right] \right\} \tag{P$^{\text{SIM}}$}$$
$$\text{s.t. (IC)}, \text{(IR)},$$
$$\mathbb{E}_{r_a} \left[U_a^{\boldsymbol{\mu}}(r_a) \right] \geqslant \mathbb{E}_{r_a} \left[U_a^{\text{BASE}}(r_a) \right] \ \ \forall a \in \mathcal{A}.$$

Using the classical approach, problem (P$^{\text{SIM}}$) reduces to the following optimization problem:

$$\max_{\boldsymbol{\mu}} \left\{ \mathbb{E}_{\mathbf{r}} \left[\sum_{a=1}^{A} \min \left\{ r_a \Theta_a^{\boldsymbol{\mu}}(\mathbf{r}), U_a^{\text{BASE}}(r_a) + \psi_a(r_a) \Theta_a^{\boldsymbol{\mu}}(\mathbf{r}) \right\} \right] \right\} - \sum_{a=1}^{A} \mathbb{E}_{r_a} \left[U_a^{\text{BASE}}(r_a) \right].$$

In the optimization problem above, the first argument inside the min $\{\cdot,\cdot\}$ expression can be analyzed in isolation through pointwise maximization of a dynamic program. However, the second argument is non-linear for a given bid vector and, when combined with the first argument, renders the optimization problem difficult to solve. The study of optimal and/or approximate mechanisms for this problem can be a useful direction for future work. ■

We now obtain a mechanism – characterized by a pair of functions $(\mathbf{\Pi}^{\text{MB}}, \mathbf{M}^{\text{MB}})$ – that is an optimal solution to problem (P^{MB}). An upper bound on the objective function of problem (P^{MB}) is easy to obtain: Since $U_a^{\boldsymbol{\mu}}(r_a) = r_a\theta_a^{\boldsymbol{\mu}}(r_a) - m_a^{\boldsymbol{\mu}}(r_a)$, the constraints in (13) are the same as:

$$\mathbb{E}_{r_a}\left[m_a^{\boldsymbol{\mu}}(r_a)\right] \leqslant \mathbb{E}_{r_a}\left[r_a\theta_a^{\boldsymbol{\mu}}(r_a)\right] - \mathbb{E}_{r_a}\left[U_a^{\text{BASE}}(r_a)\right] \ \ \forall a \in \mathcal{A}.$$

Summing these inequalities over all $a \in \mathcal{A}$, we have

$$\begin{aligned}\sum_{a=1}^{A}\mathbb{E}_{r_a}\left[m_a^{\boldsymbol{\mu}}(r_a)\right] &\leqslant \sum_{a=1}^{A}\mathbb{E}_{r_a}\left[r_a\theta_a^{\boldsymbol{\mu}}(r_a)\right] - \sum_{a=1}^{A}\mathbb{E}_{r_a}\left[U_a^{\text{BASE}}(r_a)\right]\\ &\leqslant \sum_{a=1}^{A}\mathbb{E}_{r_a}\left[r_a\theta_a^{\text{VCG}}(r_a)\right] - \sum_{a=1}^{A}\mathbb{E}_{r_a}\left[U_a^{\text{BASE}}(r_a)\right].\end{aligned}$$

The last inequality follows from the fact that the VCG allocation maximizes the social surplus. We now propose a mechanism that is feasible for problem (P^{MB}) and achieves this upper bound, and is therefore optimal for that problem. Consider the following mechanism:

$$\mathbf{\Pi}^{\text{OPT-MB}}(\mathbf{r}) = \mathbf{\Pi}^{\text{VCG}}(\mathbf{r}), \tag{14}$$

$$M_a^{\text{OPT-MB}}(\mathbf{r}) = M_a^{\text{VCG}}(\mathbf{r}) + \mathbb{E}_{r_a}\left[U_a^{\text{VCG}}(r_a)\right] - \mathbb{E}_{r_a}\left[U_a^{\text{BASE}}(r_a)\right]. \tag{15}$$

From (15), we have

$$\mathbb{E}_{\mathbf{r}}\left[M_a^{\text{OPT-MB}}(\mathbf{r})\right] = \underbrace{\mathbb{E}_{\mathbf{r}}\left[M_a^{\text{VCG}}(\mathbf{r})\right] + \mathbb{E}_{r_a}\left[U_a^{\text{VCG}}(r_a)\right]}_{=\mathbb{E}_{r_a}\left[r_a\theta_a^{\text{VCG}}(r_a)\right]} - \mathbb{E}_{r_a}\left[U_a^{\text{BASE}}(r_a)\right] \ \ \forall a \in \mathcal{A}.$$

Thus, the constraints in (13) are satisfied at equality, thereby achieving the upper bound above. Moreover, since the allocation and payments are as per VCG (up to a constant), the mechanism $(\mathbf{\Pi}^{\text{OPT-MB}}, \mathbf{M}^{\text{OPT-MB}})$ satisfies the (IC) constraints. Thus, we have

THEOREM 5. *The mechanism,* $(\mathbf{\Pi}^{\text{OPT-MB}}, \mathbf{M}^{\text{OPT-MB}})$*, is an optimal solution to problem* (P^{MB}).

We briefly note some salient features of the OPT-MB mechanism:

- The allocation rule, being VCG, is *first-best* and achieves the maximum social welfare from the sale of an impression. As discussed earlier (see (5); Section 3.3), the allocations can be computed efficiently (specifically, in time $\mathcal{O}(AN)$) by solving a dynamic program, thus making them amenable to real-time bidding.

- The payment from an advertiser is equal to the value he obtains from the allocation less his expected BASE-utility. Consequently, the total expected utility of advertiser a in the long run is equal to $\mathbb{E}_{r_a}\left[U_a^{\text{BASE}}(r_a)\right]$. Thus, the ad-exchange ensures that each advertiser obtains exactly his expected BASE-utility and keeps the remainder of the social surplus.

9.1. Applying the OPT-MB Mechanism to the Win-Win and Win-Lose Scenarios

We revisit the win-win and win-lose scenarios discussed in Section 8.1 and 8.2, respectively, and contrast the performance of the OPT-MB mechanism with that of the OPT-IR mechanism. Tables 11 and 12 provide the details of this comparison.

Win-Win Scenario (Section 8.1)					
	BASE	OPT-IR	OPT-MB	Gain in OPT-IR (w.r.t. BASE)	Gain in OPT-MB (w.r.t. BASE)
Expected Utility of X	$\frac{p}{6}$	$\frac{7p}{24}$	$\frac{p}{6}$	+75%	0%
Expected Utility of Y	$\frac{p}{6}$	$\frac{7p}{24}$	$\frac{p}{6}$	+75%	0%
Expected Revenue of Ad-Exchange	$\frac{4p}{3}$	$\frac{19p}{12}$	$\frac{83p}{12}$	+18.75%	+159.38%

Table 11 **Comparison of expected utilities of the advertisers and expected revenue of the ad exchange under the OPT-IR and the OPT-MB mechanisms under the win-win scenario (discussed in Section 8.1).**

In the win-win scenario, recall from Section 8.1 that the advertisers and the ad-exchange are better off under the OPT-IR mechanism as compared to the BASE mechanism: both the advertisers enjoy a 75% increase in their expected utility and the ad-exchange's expected revenue increases by 18.75%. Under the OPT-MB mechanism, the ad-exchange's expected revenue increases by 159.38% relative to the BASE mechanism. Thus, the ad exchange benefits more under the OPT-MB mechanism as compared to the OPT-IR mechanism. This is driven by the fact that the OPT-MB mechanism only needs to provide the expected BASE-utility to the advertisers.

Win-Lose Scenario (Section 8.2)					
	BASE	OPT-IR	OPT-MB	Gain in OPT-IR (w.r.t. BASE)	Gain in OPT-MB (w.r.t. BASE)
Expected Utility of X	$\frac{1}{8}$	$\frac{1-q}{8}$	$\frac{1}{8}$	$-100q$%	0%
Expected Utility of Y	0	0	0	0%	0%
Expected Revenue of Ad-Exchange	$\frac{3}{4}$	$\frac{3(1+q)}{4}$	$\frac{6+5q}{8}$	$+100q$%	$+83.33q$%

Table 12 **Comparison of expected utilities of the advertisers and expected revenue of the ad exchange under the OPT-IR and the OPT-MB mechanisms under the win-lose scenario (discussed in Section 8.2).**

In the win-lose scenario, advertiser X takes a hit ($-100q\%$) in his expected utility in the OPT-IR mechanism relative to the BASE mechanism, while the OPT-MB mechanism prevents this by giving advertiser X his expected BASE-utility. This guarantee that the OPT-MB mechanism provides to the advertisers comes at the expense of a reduced gain in the expected revenue ($83.33q\%$ as compared to $100q\%$) of the ad-exchange.

To summarize, the advertisers pay a price – foregoing the gain (in their respective expected utilities) they could have received in win-win scenarios – to insure themselves against any loss in the gain (relative to the base mechanism) that can arise in win-lose scenarios. Thus, neither the OPT-IR nor the OPT-MB mechanism dominates the other in the sense that the advertisers and the exchange prefer different mechanisms under different conditions.

It is encouraging to see that, despite guaranteeing the BASE-utility to advertisers, the OPT-MB mechanism benefits the ad-exchange handsomely relative to the BASE mechanism. While our purpose above was to analytically establish the tradeoff between the OPT-IR and OPT-MB mechanisms vis-à-vis the win-win and win-lose scenarios, we now inspect the performance of these mechanisms on a more-realistic test bed.

10. Numerical Investigation

Let the gain in the expected revenue of the ad-exchange under mechanism μ, relative to the BASE mechanism, be defined as follows:

$$\rho^{\mu} := \frac{\text{Expected revenue to the ad-exchange under mechanism } \mu}{\text{Expected revenue to the ad-exchange under the BASE mechanism}} - 1.$$

Similarly, let the gain in the total expected utility over *all* the advertisers under mechanism μ, relative to the BASE mechanism, be defined as follows:

$$\eta^{\mu} := \frac{\text{Total expected utility of the advertisers under mechanism } \mu}{\text{Total expected utility of the advertisers under the BASE mechanism}} - 1.$$

In our discussion below, we will refer to ρ^{μ} and η^{μ} as, respectively, the revenue gain (of the ad-exchange) and the utility gain (of the advertisers) under mechanism μ. Informed by current practice, we now describe a realistic test-bed for our study.

Test Bed

We set the number of advertisers (who are typically demand-side networks), A, to 10. Given that (i) for a majority of ads on mobile apps, an impression lasts for a maximum of 300 seconds (MarketingLand.com 2014) and (ii) the minimum duration for the display of an ad is 30 seconds at major ad-exchnages such as DoubleClick and OpenX, we set the number of time slots, N, to 10. The valuation-per-click bids of the advertisers are independently drawn from $U(0,1)$. The conditional probability, λ, that a user stays on the app at the end of a time slot, given she enters that time slot and does not click on the ad, is assumed to be 0.9.

We assume that the click-probability of each ad decays at a constant (ad-specific) rate with the passage of time, i.e., $p_{a,n+1} = \delta_a p_{a,n}$ for all $a \in \mathcal{A}$ and $n \geqslant 1$. We refer to δ_a as the decay rate[7] of ad a. Thus, under this constant decay-rate structure, the click-probabilities of an ad in any time slot are determined by its click probability in the first slot and its decay rate. Recall from Section 1.1 that the allocative inefficiency that we seek to eliminate is primarily driven by the variation in the click-probability of the ads over time. To examine the effect of this variation on the revenue gain (of the ad-exchange) and the utility gain (of the advertisers) under the OPT-IR and the OPT-MB mechanisms, we consider the following two scenarios:

- *Homogeneous ads*: In this scenario, for each ad, the click-probability in the first time slot is chosen from a common distribution. Similarly, the respective decay rates of the ads are chosen from a common distribution. Using these values, the click-probabilities of the ads in the other slots are computed. The click-through rate for mobile display ads across different formats is about 0.05% on average (SmartInsights.com 2018). Accordingly, the click-probabilities of the ads in the first time slot are independently drawn from $U(0,0.001)$. The decay rate of each ad is independently drawn from $U(0,1)$.

- *Heterogeneous ads*: In this scenario, we draw 10 random variables from $U(0,0.001)$ and sort them in ascending order – these are the click-probabilities in the first time slot of ads 1 through 10, in that order. The decay rates in the click-probabilities of the ads are drawn from $U(0,1)$ and then sorted in descending order. Thus, ads with a higher index experience a relatively sharper decay than those with a lower index. In this sense, ads with a higher index can be viewed as *impulse* ads while those with a lower index as *steady* ads.

We compute the percentage gain in the ad-exchange's revenue and the advertisers' total utility under the OPT-IR and OPT-MB mechanisms (relative to the BASE mechanism) using the sample averages, over 1,000,000 instances. Table 13 summarizes this performance under the homogeneous-ads and the heterogeneous-ads scenarios. We discuss the observations below:

[7] The OPT-IR and the OPT-MB mechanisms do not require a specific structure on the click-probabilities of the ads in different slots. We assume a constant decay-rate structure only for our numerical investigation; see Sun et al. (2017) for the mathematical micro-foundation of the diminishing click-probability of display ads with time and MarketingLand.com (2014) for empirical evidence.

	Homogeneous Ads		Heterogeneous Ads	
	OPT-IR	OPT-MB	OPT-IR	OPT-MB
Revenue Gain (ρ^μ)	7.01%	23.02%	18.22%	33.60%
Utility Gain (η^μ)	13.72%	$\approx$ 0%	55.61%	$\approx$ 0%

Table 13 **Revenue Gain and Utility Gain under the OPT-IR and the OPT-MB mechanisms for the homogeneous-ads and the heterogeneous-ads scenarios.**

- Under both the scenarios, the ad-exchange benefits handsomely under the OPT-IR as well as the OPT-MB mechanism, relative to the BASE mechanism. Thus, the partitioned selling of an impression is a promising idea to reduce the allocative inefficiency that we discussed earlier.

  In the homogeneous-ads scenario, both the ad-exchange and the advertisers are better-off under the OPT-IR mechanism (resp., by 7.01% and 13.72%). Under the OPT-MB mechanism, the revenue gain of the ad-exchange significantly improves (from 7.01% to 23.02%) – this improvement results from the fact that under the OPT-MB mechanism, the advertisers receive only their respective BASE-utilities (thus, $\eta^{\text{OPT-MB}} \approx 0\%$) and the remainder of the total welfare generated goes to the ad-exchange. The same phenomenon is observed in the heterogeneous-ads scenario, with the difference being that both the revenue gain of the ad-exchange (under the OPT-IR and the OPT-MB mechanisms) and the utility gain of the advertisers (under the OPT-IR mechanism) are more pronounced. When the ads are heteroegenous, the click-probabilities of *impulse* ads decay more sharply than those of *steady* ads, resulting in a higher allocative inefficiency as compared to the homogeneous-ads scenario. Both the mechanisms exploit this opportunity and offer higher benefits.

- In this numerical study, it turned out that under the OPT-IR mechanism, both the ad-exchange and the advertisers benefit relative to the BASE mechanism. That is, we are under the win-win scenario (Section 8.1). Note, however, that there is no guarantee offered to the advertisers in the OPT-IR mechanism. As we demonstrated in Section 8.2, it is also possible for some or all the advertisers to be worse-off under the OPT-IR mechanism relative to the BASE mechanism. On the other hand, the use of the OPT-MB mechanism allows the ad-exchange to guarantee advertisers at least their BASE-utility. Despite offering this guarantee, in our numerical study, the ad-exchange is better-off using the OPT-MB mechanism over the OPT-IR mechanism. Indeed, this is always the case in a win-win scenario. Thus, with the OPT-MB mechanism, the ad-exchange can achieve two goals simultaneously under a win-win scenario: assure advertisers of their welfare *and* obtain a higher revenue. However, in general, it is possible that in adopting the OPT-MB mechanism, the ad-exchange has to sacrifice some of its benefit to guarantee the BASE-utility to advertisers. Ultimately, the choice between the OPT-IR and the OPT-MB mechanisms will depend on the environment in which the ad-exchange operates and whether the exchange wants to assure advertisers of their long-term welfare over the BASE mechanism.

We further investigate the impact of the heterogeneity in the click-probabilities of the ads and the number of advertisers on the revenue gain of the ad-exchange. To this end, we consider: (i) six different distributions of the decay rate by systematically varying its support (see Table 14) and (ii) three values of $A: 10, 20, 30$. The values of all the other parameters are the same as before.

Distribution of Decay Rate of Ads (δ_a)	Coefficient of Variation of δ_a
$\text{Prob}(\delta_a = 0.5) = 1\ \forall a$	0
$U(0.4, 0.9)$	0.12
$U(0.3, 0.9)$	0.23
$U(0.2, 0.9)$	0.35
$U(0.1, 0.9)$	0.46
$U(0, 1)$	0.58

Table 14 Distributions of the decay rates of the click-probabilities of the ads and the corresponding coefficient of variation.

Figures 3a and 3b illustrate the impact of the variation in the decay rates of the ads on the revenue gain of the ad-exchange under, respectively, the OPT-IR and the OPT-MB mechanisms. Also shown is the change in this impact as the number of advertisers varies. The revenue gain under both the mechanisms increases with a higher variation in the decay rate. The intuition behind this is as follows. Under a small coefficient of variation of the decay rate, the click-probabilities of the ads decay at nearly the same rate. Thus, to begin with, there is little allocative inefficiency to exploit in the BASE mechanism. As the coefficient of variation of the decay rate increases, the contrast between steady and impulse ads becomes prominent, leading to an increase in the allocative inefficiency in the BASE mechanism, which the OPT-IR and the OPT-MB mechanisms duly harness. For a fixed coefficient of variation of the decay rate, an increase in the number of advertisers naturally leads to higher competition and, hence, higher revenue gains under both the mechanisms. Furthermore, the marginal revenue gain from an additional advertiser increases with the variation in the decay rate. In summary, the benefit offered by these two mechanisms (over the BASE mechanism) increases as the ads become more diverse (in their click-probability decay) and as competition for the impression increases.

We end this section by commenting on the social welfare achieved by the OPT-IR and the OPT-MB mechanisms. Although these mechanisms are designed from the perspective of the ad-exchange, they also significantly improve the social welfare relative to the BASE mechanism – Table 15 compares the

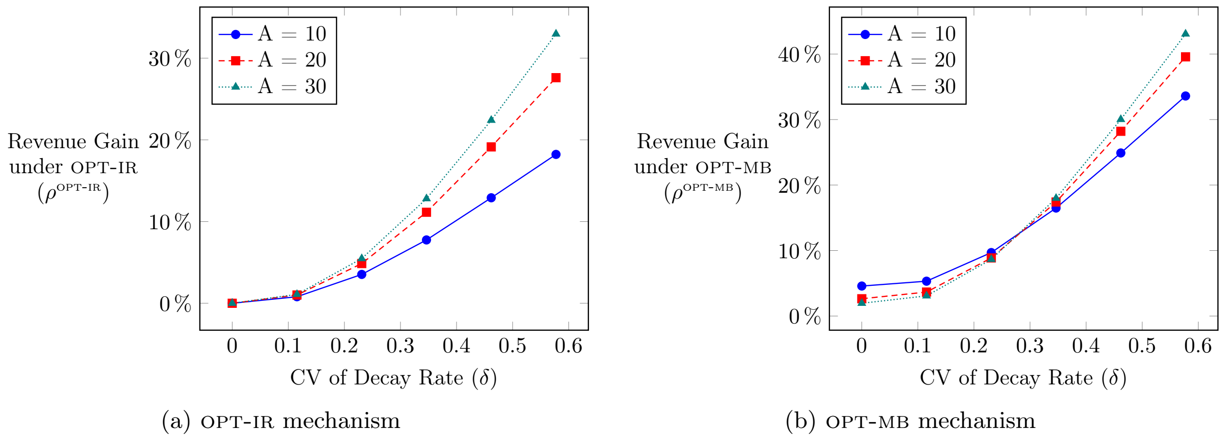


Figure 3 Revenue gain of the ad-exchange for varying distributions of the decay rate of ads and different number of advertisers under (a) OPT-IR mechanism and (b) OPT-MB mechanism.

social welfare under the three mechanisms. The BASE mechanism suffers from two distinct sources of inefficiency[8]: one, the ad-exchange retains the impression if the highest virtual valuation is negative and two, contingent on the highest virtual valuation being non-negative, a *single* ad is allocated to all the time slots. The BASE mechanism achieves 84.48% (resp., 77.45%) of the first-best social welfare under the homogeneous-ads (resp., heterogenous-ads) scenario. The OPT-IR mechanism removes the inefficiency that arises from allocating a single ad to the impression, thereby improving social welfare. Further, since this inefficiency is more pronounced in the heterogenous-ads scenario, the improvement in social welfare is higher in that scenario. The OPT-MB mechanism removes both the inefficiencies, thereby achieving the first-best social welfare.

Mechanism	Homogeneous Ads	Heterogeneous Ads
BASE	84.48%	77.45%
OPT-IR	91.54%	95.46%
OPT-MB	100%	100%

Table 15 Percentage of the first-best social welfare achieved by the BASE, OPT-IR, and OPT-MB mechanisms, under the homogeneous-ads and the heterogeneous-ads scenarios.

[8] Another source of inefficiency that could arise in a revenue-maximizing mechanism is when the impression is allocated to the advertiser with the highest *virtual* valuation instead of the advertiser with the highest valuation. However, in our numerical experiments, the advertisers' valuations are symmetric and regular, and thus our mechanisms do not suffer from this type of inefficiency.

11. Concluding Remarks

With the rapid evolution of technology in the online advertising ecosystem, it has become imperative for stakeholders to come up with solutions that improve their margins as well as the efficiency of the supply chain. The process of delivering digital ads suffers from several inefficiencies due to a fragmented supply chain. In this paper, we analyzed one such inefficiency, namely the one associated with selling an impression at once to a single advertiser, and obtained mechanisms that are amenable to the real-time bidding protocol. We hope that these mechanisms receive significant attention from the industry as they better match advertisers and publishers, and possess attractive revenue implications for ad-exchanges.

In our analysis, we have assumed that the private information (valuation-per-click r_a of advertiser a; $a \in \mathcal{A}$) of the advertisers remains unchanged during the current session. In general, the valuation-per-click of the advertisers can evolve with the passage of time in the following sense: Given that a user has not clicked on an ad displayed in the sequence thus far, the valuation-per-click of the advertisers can change in the current time slot to reflect this fact. The modeling framework that incorporates such a feature is that of *dynamic* mechanism-design, in which the private information itself changes over time. The design of dynamic mechanisms specific to our context that accommodate the changes in the private information of the advertisers is a complex problem and is worthy of future investigation.

A natural design question that arises in selling a partitioned impression is the determination of the lengths of the time slots. As of now, ad-exchanges such as OpenX and DoubleClick allow publishers to control the refresh rate of ads and offer broad guidelines for better performance. A separate study is required to understand the ad-exchange's problem of dynamically determining the lengths of time slots and estimating the click probabilities of ads in these slots.

Our analysis in this paper can potentially be applied to a variety of other settings. Stated in more-general terms, our problem of selling partitioned impressions has the following features: (i) The good that the principle allocates to the agents is "divisible" and there are "dependencies" among the divided sub-goods (i.e., the divided sub-goods are "connected" to each other) and (ii) The valuations of the divided sub-goods are heterogeneous (i.e., the agents value them differently). Similarly, our objective in more-general terms is to obtain a "win-win" solution for both the principal and the agents, vis-à-vis the current practice. Two examples of other potential applications are as follows:

- The use of shared resources (such as computing capacity) by multiple agents. Here, the division of the good is represented by the allocation of time for which the resources are assigned to different agents. Such a setting offers several interesting constraints on the allocation of the resources. For example, some agents may be endowed with non-preemptive jobs, which may necessitate the designer to allocate only adjacent slots to these agents. Further, there may be different tiers/qualities of the resources (e.g., high-, medium-, low-computing capacity) and agents may require specific proportions of each of these resources.

- The use of a shared space by multiple agents. An example of such a setting is the allocation of real-estate to multiple retail outlets in a shopping mall. Here, the division of the good (real-estate) is represented by the percentage allocation of the shared space to different retail outlets. In this case, the physical space (area as well as location) allocated to one outlet might affect customer demand at other outlets.

Given the competitive nature of the online advertising landscape, it is natural for the industry to continually look for better ways to monetize the ad-delivery process. For instance, to overcome the inefficiency in discovering the best price available for an impression across different ad-networks and ad-exchanges, the industry has started adopting the *header-bidding technology* to enable publishers to consolidate bids across different ad-networks and exchanges, and essentially run a simultaneous auction locally on the user's app/website.[9] The header-bidding mechanism and the mechanisms that we discussed in this paper help improve, respectively, the demand-side and the supply-side market thickness. In future research, it would be interesting to analyze the interplay of these two mechanisms and the improvement they offer to the ecosystem in conjunction.

References

Abhishek, V., and K. Hosanagar. 2013. Optimal bidding in multi-item multislot sponsored search auctions. *Operations Research* 61 (4): 855–873.

Adaptly 2014. A research study on sequenced for call to action vs. sustained call to action. [Accessed: July, 2019] `http://adaptly.com/wp-content/uploads/2014/11/Adaptly-Refinery29-White-Paper-2014.pdf`.

Agarwal, A., K. Hosanagar, and M. D. Smith. 2011. Location, location, location: An analysis of profitability of position in online advertising markets. *Journal of Marketing Research* 48 (6): 1057–1073.

Allouah, A., and O. Besbes. 2017. Auctions in the online display advertising chain: A case for independent campaign management. Available at SSRN: `https://ssrn.com/abstract=2919665`.

Andrews, M., X. Luo, Z. Fang, and A. Ghose. 2015. Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Marketing Science* 35 (2): 218–233.

[9] For more details, see https://martechtoday.com/martech-landscape-what-is-header-bidding-and-why-should-publishers-care-157065.

Asdemir, K., N. Kumar, and V. S. Jacob. 2012. Pricing models for online advertising: CPM vs. CPC. *Information Systems Research* 23 (3-part-1): 804–822.

Aseri, M., M. Dawande, G. Janakiraman, and V. Mookerjee. 2017. Procurement policies for mobile-promotion platforms. *Management Science* 64 (10): 4590–4607.

Aumann, Y., Y. Dombb, and A. Hassidim. 2016. Auctioning time: Truthful auctions of heterogeneous divisible goods. *ACM Transactions on Economics and Computation* 4 (1): 3.

Balseiro, S., A. Kim, M. Mahdian, and V. Mirrokni. 2017. Budget management strategies in repeated auctions. In *Proceedings of the 26th International Conference on World Wide Web*, 15–23. International World Wide Web Conferences Steering Committee.

Balseiro, S. R., O. Besbes, and G. Y. Weintraub. 2015. Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science* 61 (4): 864–884.

Balseiro, S. R., J. Feldman, V. Mirrokni, and S. Muthukrishnan. 2014. Yield optimization of display advertising with ad exchange. *Management Science* 60 (12): 2886–2907.

Bharadwaj, V., P. Chen, W. Ma, C. Nagarajan, J. Tomlin, S. Vassilvitskii, E. Vee, and J. Yang. 2012. SHALE:An efficient algorithm for allocation of guaranteed display advertising. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1195–1203. ACM.

Braun, M., and W. W. Moe. 2013. Online display advertising: Modeling the effects of multiple creatives and individual impression histories. *Marketing Science* 32 (5): 753–767.

Chatterjee, P., D. L. Hoffman, and T. P. Novak. 2003. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science* 22 (4): 520–541.

Chen, Y., X. Li, and M. Sun. 2017. Competitive mobile geo targeting. *Marketing Science* 36 (5): 666–682.

Edelman, B., M. Ostrovsky, and M. Schwarz. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review* 97 (1): 242–259.

Garg, D., and Y. Narahari. 2009. An optimal mechanism for sponsored search auctions on the web and comparison with other mechanisms. *IEEE Transactions on Automation Science and Engineering* 6 (4): 641–657.

Goldfarb, A., and C. Tucker. 2011. Online display advertising: Targeting and obtrusiveness. *Marketing Science* 30 (3): 389–404.

Goldstein, D. G., R. P. McAfee, and S. Suri. 2015. Improving the effectiveness of time-based display advertising. *ACM Transactions on Economics and Computation* 3 (2): 7.

Hojjat, A., J. Turner, S. Cetintas, and J. Yang. 2017. A unified framework for the scheduling of guaranteed targeted display advertising under reach and frequency requirements. *Operations Research* 65 (2): 289–313.

iAd 2014. Inside mobile advertising: top 5 trends and insights. [Accessed: July, 2019] `http://www.callmemobi.com/wp-content/uploads/2013/04/MCSAATCHIMOBILEv15.pdf`.

Interactive Advertising Bureau 2017. IAB Internet advertising revenue report. [Accessed: July, 2019] `https://www.iab.com/wp-content/uploads/2018/05/IAB-2017-Full-Year-Internet-Advertising-Revenue-Webinar-Presentation.pdf`.

Kim, A., S. Balachander, and K. Kannan. 2012. On the optimal number of advertising slots in a generalized second-price auction. *Marketing Letters* 23 (3): 851–868.

Korula, N., V. Mirrokni, and H. Nazerzadeh. 2016. Optimizing display advertising markets: Challenges and directions. *IEEE Internet Computing* 20 (1): 28–35.

Krishna, V. 2009. *Auction Theory*. Academic press.

Krishna, V., and M. Perry. 1998. Efficient mechanism design. Available at SSRN: `https://ssrn.com/abstract=64934`.

Liu, D., J. Chen, and A. B. Whinston. 2010. Ex ante information and the design of keyword auctions. *Information Systems Research* 21 (1): 133–153.

Mansour, Y., S. Muthukrishnan, and N. Nisan. 2012. Doubleclick ad exchange auction. *arXiv preprint arXiv:1204.0535*.

MarketingLand.com 2014, July. The shelf life of a mobile ad: Shorter than you may think. [Accessed: July, 2019] `http://marketingland.com/shelf-life-mobile-ad-shorter-may-think-91495`.

McAfee, R. P., and S. Vassilvitskii. 2012. An overview of practical exchange design. *Current Science*:1056–1063.

Mohan, P., S. Nath, and O. Riva. 2013. Prefetching mobile ads: Can advertising systems afford it? In *Proceedings of the 8th ACM European Conference on Computer Systems*, 267–280. ACM.

Muthukrishnan, S. 2009. Ad exchanges: Research issues. In *International Workshop on Internet and Network Economics*, 1–12. Springer.

Myerson, R. B. 1981. Optimal auction design. *Mathematics of Operations Research* 6 (1): 58–73.

Najafi-Asadolahi, S., and K. Fridgeirsdottir. 2014. Cost-per-click pricing for display advertising. *Manufacturing & Service Operations Management* 16 (4): 482–497.

Nisan, N., T. Roughgarden, E. Tardos, and V. V. Vazirani. 2007. *Algorithmic game theory*, Volume 1. Cambridge University Press Cambridge.

Roels, G., and K. Fridgeirsdottir. 2009. Dynamic revenue management for online display advertising. *Journal of Revenue and Pricing Management* 8 (5): 452–466.

SmartInsights.com 2018, August. Average display advertising clickthrough rates. [Accessed: July, 2019] `https://www.smartinsights.com/internet-advertising/internet-advertising-analytics/display-advertising-clickthrough-rates/`.

Sun, Z., M. Dawande, G. Janakiraman, and V. Mookerjee. 2017. Not just a fad: Optimal sequencing in mobile in-app advertising. *Information Systems Research* 28 (3): 511–528.

Thompson, D. R., and K. Leyton-Brown. 2013. Revenue optimization in the generalized second-price auction. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, 837–852. ACM.

Turner, J., A. Scheller-Wolf, and S. Tayur. 2011. OR PRACTICE – Scheduling of dynamic in-game advertising. *Operations Research* 59 (1): 1–16.

Varian, H. R. 2007. Position auctions. *International Journal of Industrial Organization* 25 (6): 1163–1178.

Yang, J., E. Vee, S. Vassilvitskii, J. Tomlin, J. Shanmugasundaram, T. Anastasakos, and O. Kennedy. 2010. Inventory Allocation for Online Graphical Display Advertising. *CoRR* abs/1008.3551.

Yuan, Y., F. Wang, J. Li, and R. Qin. 2014. A survey on real time bidding advertising. In *IEEE International Conference on Service Operations and Logistics, and Informatics*, 418–423. IEEE.

Zhang, W., S. Yuan, and J. 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. ACM.

Online Appendices to **Sustaining a Good Impression: Mechanisms for Selling 'Partitioned' Impressions at Ad-Exchanges**

Appendix A Proof of Theorem 1

Our goal is to find an optimal mechanism $(\mathbf{\Pi}^{\boldsymbol{\mu}}, \mathbf{M}^{\boldsymbol{\mu}})$ for selling partitioned impressions that is both incentive compatible and individually rational. Our analysis below follows the technique that is used to derive optimal mechanisms where agents are endowed with single-dimensional private information (Myerson 1981).

We recall the following notation: (i) The private valuation-per-click for advertiser a is denoted by r_a, which is a random variable independently distributed with publicly-known c.d.f. $F_a(\cdot)$ and p.d.f. $f_a(\cdot)$ over the interval $\mathcal{B}_a = [0, \omega_a]$. (ii) $\boldsymbol{\mathcal{B}} = \times_{a=1}^{A} \mathcal{B}_a$ denotes the Cartesian product of the valuation intervals of the advertisers. (iii) $\theta_a^{\boldsymbol{\mu}}(b_a)$ denotes the expected click-probability of ad a across all the slots, under mechanism $\boldsymbol{\mu}$, when advertiser a bids b_a and all the other advertisers bid their true valuations. (iv) $m_a^{\boldsymbol{\mu}}(b_a)$ denotes the expected payment by advertiser a to the ad-exchange, under mechanism $\boldsymbol{\mu}$, when he bids b_a and all the other advertisers report their true valuations. (v) $U_a^{\boldsymbol{\mu}}(b_a; r_a) = r_a \theta_a^{\boldsymbol{\mu}}(b_a) - m_a^{\boldsymbol{\mu}}(b_a)$ denotes the net utility to advertiser a, under mechanism $\boldsymbol{\mu}$, when he bids b_a and his true valuation-per-click is r_a.

A mechanism $(\mathbf{\Pi}^{\boldsymbol{\mu}}, \mathbf{M}^{\boldsymbol{\mu}})$ is incentive compatible if and only if $U_a^{\boldsymbol{\mu}}(r_a; r_a) \geqslant U_a^{\boldsymbol{\mu}}(b_a; r_a)$, for all $a \in \mathcal{A}$ and $r_a, b_a \in \mathcal{B}_a$. Let $U_a^{\boldsymbol{\mu}}(r_a) := U_a^{\boldsymbol{\mu}}(r_a; r_a)$. Then, the mechanism $\boldsymbol{\mu}$ is incentive compatible for the advertisers iff:

$$U_a^{\boldsymbol{\mu}}(r_a) = \max_{b_a \in \mathcal{B}_a} \; [r_a \theta_a^{\boldsymbol{\mu}}(b_a) - m_a^{\boldsymbol{\mu}}(b_a)] \;\forall a \in \mathcal{A}.$$

Since $U_a^{\boldsymbol{\mu}}$ is a family of affine functions, $U_a^{\boldsymbol{\mu}}(\cdot)$ is a convex function of its argument. Therefore, $\forall r_a, b_a$, we have:

$$\begin{aligned} b_a \theta_a^{\boldsymbol{\mu}}(r_a) - m_a^{\boldsymbol{\mu}}(r_a) &= r_a \theta_a^{\boldsymbol{\mu}}(r_a) - m_a^{\boldsymbol{\mu}}(r_a) + \theta_a^{\boldsymbol{\mu}}(r_a)(b_a - r_a) \\ &= U_a^{\boldsymbol{\mu}}(r_a) + \theta_a^{\boldsymbol{\mu}}(r_a)(b_a - r_a). \end{aligned}$$

Thus, incentive compatibility is equivalent to the requirement that for all r_a, b_a,

$$U_a^{\boldsymbol{\mu}}(b_a) \geqslant U_a^{\boldsymbol{\mu}}(r_a) + \theta_a^{\boldsymbol{\mu}}(r_a)(b_a - r_a).$$

Since a convex function is absolutely continuous, it is differentiable almost everywhere in the interior of its domain. Therefore,

$$\frac{dU_a^{\boldsymbol{\mu}}(r_a)}{dr_a} = \theta_a^{\boldsymbol{\mu}}(r_a).$$

Note that since $U_a^{\boldsymbol{\mu}}(\cdot)$ is a convex function of its argument; this implies that a mechanism $\boldsymbol{\mu}$ is incentive compatible if and only if the associated $\theta_a^{\boldsymbol{\mu}}(\cdot)$ is nondecreasing for all $a \in \mathcal{A}$. Next, as every absolutely continuous function is the definite integral of its derivative, we have

$$U_a^{\boldsymbol{\mu}}(r_a) = U_a^{\boldsymbol{\mu}}(0) + \int_0^{r_a} \theta_a^{\boldsymbol{\mu}}(t_a) dt_a. \tag{16}$$

Since $U_a^{\boldsymbol{\mu}}(0) = -m_a^{\boldsymbol{\mu}}(0)$, we have

$$m_a^{\boldsymbol{\mu}}(r_a) = m_a^{\boldsymbol{\mu}}(0) + r_a \theta_a^{\boldsymbol{\mu}}(r_a) - \int_0^{r_a} \theta_a^{\boldsymbol{\mu}}(t_a) dt_a \ \forall a \in \mathcal{A}. \tag{17}$$

For an incentive compatible mechanism $\boldsymbol{\mu}$ to be individually rational, we should have $U_a^{\boldsymbol{\mu}}(r_a) \geqslant 0 \ \forall a \in \mathcal{A}, r_a \in \mathcal{B}_a$. Since $\theta_a^{\boldsymbol{\mu}}(\cdot)$ is nondecreasing, then from (16), individual rationality is equivalent to the requirement that $U_a^{\boldsymbol{\mu}}(0) \geqslant 0$, and since $U_a^{\boldsymbol{\mu}}(0) = -m_a^{\boldsymbol{\mu}}(0)$, this is equivalent to the requirement that $m_a^{\boldsymbol{\mu}}(0) \leqslant 0$.

The expected revenue to the ad-exchange from using a mechanism $\boldsymbol{\mu}$ is simply $\sum_{a=1}^{A} \mathbb{E}_{r_a}\left[m_a^{\boldsymbol{\mu}}(r_a)\right]$.

$$\begin{aligned}
\sum_{a=1}^{A} \mathbb{E}_{r_a}\left[m_a^{\boldsymbol{\mu}}(r_a)\right] &= \sum_{a=1}^{A} \int_0^{\omega_a} m_a^{\boldsymbol{\mu}}(r_a) f_a(r_a) dr_a \\
&= \sum_{a=1}^{A} \int_0^{\omega_a} \left(m_a^{\boldsymbol{\mu}}(0) + r_a \theta_a^{\boldsymbol{\mu}}(r_a) - \int_0^{r_a} \theta_a^{\boldsymbol{\mu}}(t_a) dt_a \right) f_a(r_a) dr_a \\
&= \sum_{a=1}^{A} m_a^{\boldsymbol{\mu}}(0) + \sum_{a=1}^{A} \int_0^{\omega_a} r_a \theta_a^{\boldsymbol{\mu}}(r_a) f_a(r_a) dr_a - \sum_{a=1}^{A} \int_0^{\omega_a} \int_0^{r_a} \theta_a^{\boldsymbol{\mu}}(t_a) f_a(r_a) dt_a dr_a.
\end{aligned}$$

Changing order of integration in the last term, we get:

$$\begin{aligned}
\sum_{a=1}^{A} \int_0^{\omega_a} \int_0^{r_a} \theta_a^{\boldsymbol{\mu}}(t_a) f_a(r_a) dt_a dr_a &= \sum_{a=1}^{A} \int_0^{\omega_a} \int_{t_a}^{\omega_a} \theta_a^{\boldsymbol{\mu}}(t_a) f_a(r_a) dr_a dt_a \\
&= \sum_{a=1}^{A} \int_0^{\omega_a} \left(1 - F_a(t_a)\right) \theta_a^{\boldsymbol{\mu}}(t_a) dt_a.
\end{aligned}$$

Thus, we can write

$$\begin{aligned}
\sum_{a=1}^{A} \mathbb{E}_{r_a}\left[m_a^{\boldsymbol{\mu}}(r_a)\right] &= \sum_{a=1}^{A} m_a^{\boldsymbol{\mu}}(0) + \sum_{a=1}^{A} \int_0^{\omega_a} r_a \theta_a^{\boldsymbol{\mu}}(r_a) f_a(r_a) dr_a - \sum_{a=1}^{A} \int_0^{\omega_a} \left(1 - F_a(t_a)\right) \theta_a^{\boldsymbol{\mu}}(t_a) dt_a \\
&= \sum_{a=1}^{A} m_a^{\boldsymbol{\mu}}(0) + \sum_{a=1}^{A} \int_0^{\omega_a} \left(r_a - \frac{1 - F_a(r_a)}{f_a(r_a)} \right) \theta_a^{\boldsymbol{\mu}}(r_a) f_a(r_a) dr_a \\
&= \sum_{a=1}^{A} m_a^{\boldsymbol{\mu}}(0) + \sum_{a=1}^{A} \int_{\boldsymbol{\mathcal{B}}} \left(r_a - \frac{1 - F_a(r_a)}{f_a(r_a)} \right) \Theta_a^{\boldsymbol{\mu}}(\mathbf{r}) f(\mathbf{r}) d\mathbf{r}.
\end{aligned}$$

Notice that the expected revenue to the ad-exchange is increasing in $m_a^{\boldsymbol{\mu}}(0)$ $\forall a \in \mathcal{A}$. Since the individual-rationality constraints require $m_a^{\boldsymbol{\mu}}(0) \leqslant 0$ $\forall a \in \mathcal{A}$, for any given allocation rule, a revenue-maximizing payment rule will have $m_a^{\boldsymbol{\mu}}(0) = 0$. The ad-exchange's objective is therefore to find a mechanism $\boldsymbol{\mu}$ that maximizes

$$\sum_{a=1}^{A} \int_{\boldsymbol{\mathcal{B}}} \left(r_a - \frac{1 - F_a(r_a)}{f_a(r_a)} \right) \Theta_a^{\boldsymbol{\mu}}(\mathbf{r}) f(\mathbf{r}) d\mathbf{r} \tag{18}$$

subject to the constraint that the mechanism is incentive compatible, which is equivalent to the requirement that $\theta_a^{\boldsymbol{\mu}}(\cdot)$ is nondecreasing for all $a \in \mathcal{A}$ and that (16) be satisfied.

Finally, recall from Section 3.1 that $\psi_a(r_a) := r_a - \frac{1-F_a(r_a)}{f_a(r_a)}$ is the virtual valuation of advertiser a with valuation-per-click r_a. We assume that the distribution $f_a(\cdot)$ is regular for all $a \in \mathcal{A}$ (see Section 3.1), and therefore, the virtual valuation $\psi_a(r_a)$ is an increasing function in r_a. In the above formulation, each advertiser's private valuation-per-click r_a is "adjusted" to $r_a - \frac{1-F_a(r_a)}{f_a(r_a)}$. This adjustment stems from the fact that each advertiser's valuation-per-click is private knowledge (i.e., it is known only to that advertiser) and the ad-exchange only knows the distribution of these private values. The revenue maximization problem of the ad-exchange can now be written as:

$$\max_{\boldsymbol{\mu}} \sum_{a=1}^{A} \int_{\boldsymbol{\mathcal{B}}} \psi_a(r_a) \Theta_a^{\boldsymbol{\mu}}(\mathbf{r}) f(\mathbf{r}) d\mathbf{r} \tag{19}$$

subject to incentive compatibility constraints and that (16) be satisfied.

Consider the following mechanism:

$$\begin{aligned}
\boldsymbol{\Pi}^{\text{OPT-IR}}(\mathbf{r}) &= \arg\max_{\boldsymbol{\mu}} \left\{ \sum_{a=1}^{A} \psi_a(r_a) \Theta_a^{\boldsymbol{\mu}}(\mathbf{r}) \right\}, \\
M_a^{\text{OPT-IR}}(\mathbf{r}) &= r_a \Theta_a^{\text{OPT-IR}}(\mathbf{r}) - \int_0^{r_a} \Theta_a^{\text{OPT-IR}}(t_a, \mathbf{r}_{-a})\, dt_a \;\; \forall a \in \mathcal{A}.
\end{aligned}$$

Notice that (i) the allocation rule, $\boldsymbol{\Pi}^{\text{OPT-IR}}(\mathbf{r})$, is obtained by point-wise maximization of the objective function in (19), and (ii) the payment rule, $\mathbf{M}^{\text{OPT-IR}}(\mathbf{r})$, satisfies (16). Therefore, the mechanism $(\boldsymbol{\Pi}^{\text{OPT-IR}}, \mathbf{M}^{\text{OPT-IR}})$ is an optimal solution to the above problem if for every a, $\Theta_a^{\text{OPT-IR}}(\cdot, \mathbf{r}_{-a})$ is a non-decreasing function of its argument. ■

Appendix B Infinite-Horizon Mechanisms

A direct mechanism, $\boldsymbol{\mu}_\infty$, for the infinite-horizon setting consists of a pair of functions $(\boldsymbol{\Pi}^{\boldsymbol{\mu}_\infty}, \mathbf{M}^{\boldsymbol{\mu}_\infty})$, where $\boldsymbol{\Pi}^{\boldsymbol{\mu}_\infty} : \boldsymbol{\mathcal{B}} \to \{\mathcal{A} \cup \phi\}^\infty$ is the allocation rule that specifies the ad sequence and $\mathbf{M}^{\boldsymbol{\mu}_\infty} : \boldsymbol{\mathcal{B}} \to \mathbb{R}^A$ is the payment rule that specifies the expected payment by the advertisers to the ad-exchange.

Let $\mathbf{\Theta}_a^{\boldsymbol{\mu}_\infty}(\mathbf{b})$ denote the likelihood of a click on ad a over all the (infinite) slots, under the mechanism $\boldsymbol{\mu}_\infty$, when the advertisers bid $\mathbf{b}$. Formally,

$$\mathbf{\Theta}_a^{\boldsymbol{\mu}_\infty}(\mathbf{b}) = \sum_{n=1}^{\infty} p_{a,n} \mathbb{1}\left\{\pi_n^{\boldsymbol{\mu}_\infty}(\mathbf{b}) = a\right\} \lambda^{n-1} \prod_{t=1}^{n-1} (1 - p_{\pi_t^{\boldsymbol{\mu}_\infty}(\mathbf{b}),t}) \;\; \forall a \in \mathcal{A}.$$

The IC and the IR constraints can now be stated as follows:

$$r_a \theta_a^{\boldsymbol{\mu}_\infty}(r_a) - m_a^{\boldsymbol{\mu}_\infty}(r_a) \geqslant r_a \theta_a^{\boldsymbol{\mu}_\infty}(b_a) - m_a^{\boldsymbol{\mu}_\infty}(b_a) \;\; \forall a \in \mathcal{A}, \;\; \forall r_a, b_a \in \mathcal{B}_a \tag{IC-∞}$$

$$r_a \theta_a^{\boldsymbol{\mu}_\infty}(r_a) - m_a^{\boldsymbol{\mu}_\infty}(r_a) \geqslant 0 \;\; \forall a \in \mathcal{A}, \;\; \forall r_a, b_a \in \mathcal{B}_a \tag{IR-∞}$$

The infinite-horizon problem for Setting 1 can now be formulated as:

$$\max_{\boldsymbol{\mu}_\infty} \;\; \sum_{a=1}^{A} \mathbb{E}_{\mathbf{r}}\left[m_a^{\boldsymbol{\mu}_\infty}(\mathbf{r})\right], \text{s.t (IC-}\infty\text{)}, \text{(IR-}\infty\text{)}. \tag{P$_\infty^{\text{IR}}$}$$

Using the same approach as in the paper, an optimal solution to problem (P$_\infty^{\text{IR}}$), $(\mathbf{\Pi}^{\text{OPT-IR}\infty}, \mathbf{M}^{\text{OPT-IR}\infty})$, is given by:

$$\begin{aligned}
\mathbf{\Pi}^{\text{OPT-IR}\infty}(\mathbf{r}) &= \arg\max_{\boldsymbol{\mu}_\infty} \left\{ \sum_{a=1}^{A} \psi_a(r_a) \Theta_a^{\boldsymbol{\mu}_\infty}(\mathbf{r}) \right\}, \\
M_a^{\text{OPT-IR}\infty}(\mathbf{r}) &= r_a \Theta_a^{\text{OPT-IR}\infty}(\mathbf{r}) - \int_0^{r_a} \Theta_a^{\text{OPT-IR}\infty}(t_a, \mathbf{r}_{-a}) \, dt_a \;\; \forall a \in \mathcal{A}.
\end{aligned}$$

Further, the optimal revenue of the ad exchange, denoted as *Revenue*(OPT-IR$_\infty$), under the mechanism $(\mathbf{\Pi}^{\text{OPT-IR}\infty}(\mathbf{r}), \mathbf{M}^{\text{OPT-IR}\infty}(\mathbf{r}))$ is:

$$\mathit{Revenue}(\text{OPT-IR}_\infty) = \sum_{a \in \mathcal{A}^+} \psi_a(r_a) \Theta_a^{\text{OPT-IR}\infty}(\mathbf{r}).$$

Here, we note that the optimization problem that appears in the definition of (P$_\infty^{\text{IR}}$) can no longer be formulated as a finite-stage dynamic program. However, as we will show below, we do not need to solve this optimization problem explicitly in order to arrive at our conclusion.

We now turn our attention to the finite-horizon mechanisms discussed in the paper. Consistent with the notation for the infinite horizon mechanisms, we use the subscript N for a finite horizon mechanism with N time slots. Thus, a direct mechanism for selling N partitioned impressions is denoted by $\boldsymbol{\mu}_N$ and is characterized by the allocation function $\mathbf{\Pi}^{\boldsymbol{\mu}_N} : \boldsymbol{\mathcal{B}} \to \{\mathcal{A} \cup \phi\}^N$, and the payment function $\mathbf{M}^{\boldsymbol{\mu}_N} : \boldsymbol{\mathcal{B}} \to \mathbb{R}^A$. Finally, for mechanism $\boldsymbol{\mu}_N$, let $\mathbf{\Theta}_a^{\boldsymbol{\mu}_N}(\mathbf{b})$ denote the likelihood of a click on ad a over all the (infinite) slots, under the mechanism $\boldsymbol{\mu}_\infty$, when the advertisers bid $\mathbf{b}$. Thus,

$$\Theta_a^{\boldsymbol{\mu}_N}(\mathbf{r}) = \sum_{n=1}^{N} p_{a,n} \mathbb{1}\left\{\pi_n^{\boldsymbol{\mu}_N}(\mathbf{b}) = a\right\} \lambda^{n-1} \prod_{t=1}^{n-1} (1 - p_{\pi_t^{\boldsymbol{\mu}_N}(\mathbf{b}),t}) \;\; \forall a \in \mathcal{A}.$$

Recall that the optimal finite-horizon mechanism, $(\mathbf{\Pi}^{\text{OPT-IR}_N}(\mathbf{r}), \mathbf{M}^{\text{OPT-IR}_N}(\mathbf{r}))$, is given by:

$$\mathbf{\Pi}^{\text{OPT-IR}_N}(\mathbf{r}) = \arg\max_{\boldsymbol{\mu}_N} \left\{ \sum_{a=1}^{A} \psi_a(r_a)\Theta_a^{\boldsymbol{\mu}_N}(\mathbf{r}) \right\},$$
$$M_a^{\text{OPT-IR}_N}(\mathbf{r}) = r_a\Theta_a^{\text{OPT-IR}_N}(\mathbf{r}) - \int_0^{r_a} \Theta_a^{\text{OPT-IR}_N}(t_a, \mathbf{r}_{-a})\, dt_a \;\; \forall a \in \mathcal{A},$$

and the optimal revenue under $(\mathbf{\Pi}^{\text{OPT-IR}_N}(\mathbf{r}), \mathbf{M}^{\text{OPT-IR}_N}(\mathbf{r}))$ is given by:

$$\mathit{Revenue}(\text{OPT-IR}_N) = \sum_{a \in \mathcal{A}^+} \psi_a(r_a)\Theta_a^{\text{OPT-IR}_N}(\mathbf{r}). \tag{20}$$

We now show that, if $\lambda < 1$, then by choosing an appropriate length of a finite horizon, the revenue of the resulting optimal mechanism can be made arbitrarily close to that of the optimal infinite-horizon mechanism. Mathematically, we have

Proposition 1. *Consider any $\epsilon > 0$. Then, for $0 \leqslant \lambda < 1$, there exists a positive integer $N(\epsilon) < \infty$ such that:*

$$\mathit{Revenue}(\text{OPT-IR}_\infty) - \mathit{Revenue}(\text{OPT-IR}_{N(\epsilon)}) < \epsilon.$$

Proof: For any infinite-horizon mechanism $\boldsymbol{\mu}_\infty$ and a given bid vector $\mathbf{b}$, let

$$h(n; \mathbf{b}, \boldsymbol{\mu}_\infty) = \lambda^{n-1} \prod_{t=1}^{n-1} (1 - p_{\pi_t^{\boldsymbol{\mu}_\infty}(\mathbf{b}),t})$$

denote the probability that in the first $n-1$ slots, neither the user exited the app (λ^{n-1}) nor did she click on any ad $\left(\prod_{t=1}^{n-1}(1 - p_{\pi_t^{\boldsymbol{\mu}_\infty}(\mathbf{b}),t})\right)$. Consider an arbitrary $\epsilon > 0$. Let $\Delta = \frac{\sum_{a\in\mathcal{A}^+} \psi_a(r_a)}{1-\lambda}$, which is a finite and positive constant, and let $\epsilon' = \epsilon/\Delta > 0$. For $0 \leqslant \lambda < 1$, $h(n; \mathbf{b}, \boldsymbol{\mu}_\infty)$ is decreasing in n, and $\lim_{n\to\infty} h(n; \mathbf{b}, \boldsymbol{\mu}_\infty) = 0$. Therefore, there exists a positive integer $M(\epsilon')$, which can also be denoted as $N(\epsilon)$, such that $h(n; \mathbf{b}, \boldsymbol{\mu}_\infty) < \epsilon' \; \forall n > N(\epsilon)$. Further, for $k \geqslant 1$, we have:

$$\begin{aligned} h(N(\epsilon) + k; \mathbf{b}, \boldsymbol{\mu}_\infty) &= \lambda^{N(\epsilon)+k-1} \prod_{t=1}^{N(\epsilon)+k-1} (1 - p_{\pi_t^{\boldsymbol{\mu}_\infty}(\mathbf{b}),t}) \\ &= \lambda(1 - p_{\pi_{N(\epsilon)+k-1}^{\boldsymbol{\mu}_\infty}(\mathbf{b}),N(\epsilon)+k-1}) h(N(\epsilon)+k-1; \mathbf{b}, \boldsymbol{\mu}_\infty) \\ &\leqslant \lambda h(N(\epsilon)+k-1; \mathbf{b}, \boldsymbol{\mu}_\infty) \;\; (\text{since } p_{\pi_{N(\epsilon)+k-1}^{\boldsymbol{\mu}_\infty}(\mathbf{b}),N(\epsilon)+k-1} \leqslant 1) \end{aligned}$$

Using this recursion, we get:

$$h(N(\epsilon) + k; \mathbf{b}, \boldsymbol{\mu}_\infty) < \lambda^{k-1}\epsilon' \;\; \forall k \geqslant 1. \tag{21}$$

Then, we have

$$
\begin{aligned}
&\mathit{Revenue}(\textsc{opt-ir}_\infty)\\
&= \sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{n=1}^{\infty}p_{a,n}\mathbb{1}\left\{\pi_n^{\textsc{opt-ir}_\infty}(\mathbf{r})=a\right\}\lambda^{n-1}\prod_{t=1}^{n-1}(1-p_{\pi_t^{\textsc{opt-ir}_\infty}(\mathbf{r}),t})\\
&= \sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{n=1}^{\infty}p_{a,n}\mathbb{1}\left\{\pi_n^{\textsc{opt-ir}_\infty}(\mathbf{r})=a\right\}h(n;\mathbf{r},\textsc{opt-ir}_\infty)\\
&= \sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{n=1}^{N(\epsilon)}p_{a,n}\mathbb{1}\left\{\pi_n^{\textsc{opt-ir}_\infty}(\mathbf{r})=a\right\}h(n;\mathbf{r},\textsc{opt-ir}_\infty)\ +\\
&\sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{n=N(\epsilon)+1}^{\infty}p_{a,n}\mathbb{1}\left\{\pi_n^{\textsc{opt-ir}_\infty}(\mathbf{r})=a\right\}h(n;\mathbf{r},\textsc{opt-ir}_\infty)\\
&= \sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{n=1}^{N(\epsilon)}p_{a,n}\mathbb{1}\left\{\pi_n^{\textsc{opt-ir}_\infty}(\mathbf{r})=a\right\}h(n;\mathbf{r},\textsc{opt-ir}_\infty)\ +\\
&\sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{k=1}^{\infty}p_{a,N(\epsilon)+k}\mathbb{1}\left\{\pi_{N(\epsilon)+k}^{\boldsymbol{\mu}}(\mathbf{r})=a\right\}h(N(\epsilon)+k;\mathbf{r},\textsc{opt-ir}_\infty)\\
&< \sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{n=1}^{N(\epsilon)}p_{a,n}\mathbb{1}\left\{\pi_n^{\textsc{opt-ir}_\infty}(\mathbf{r})=a\right\}h(n;\mathbf{r},\textsc{opt-ir}_\infty)+\sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{k=1}^{\infty}\lambda^{k-1}\epsilon'\ \text{(from (21))}\\
&\leqslant \sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{n=1}^{N(\epsilon)}p_{a,n}\mathbb{1}\left\{\pi_n^{\textsc{opt-ir}_\infty}(\mathbf{r})=a\right\}h(n;\mathbf{r},\textsc{opt-ir}_\infty)+\frac{\epsilon'\cdot\sum_{a\in\mathcal{A}^+}\psi_a(r_a)}{1-\lambda}\ \text{(since } 0\leqslant\lambda<1)\\
&\leqslant \sum_{a\in\mathcal{A}^+}\psi_a(r_a)\cdot\sum_{n=1}^{N(\epsilon)}p_{a,n}\mathbb{1}\left\{\pi_n^{\textsc{opt-ir}_{N(\epsilon)}}(\mathbf{r})=a\right\}h(n;\mathbf{r},\textsc{opt-ir}_{N(\epsilon)})+\frac{\epsilon'\cdot\sum_{a\in\mathcal{A}^+}\psi_a(r_a)}{1-\lambda}\ \text{(definition of } \textsc{opt-ir}_{N(\epsilon)})\\
&\leqslant \mathit{Revenue}(\textsc{opt-ir}_{N(\epsilon)})+\frac{\epsilon'\cdot\sum_{a\in\mathcal{A}^+}\psi_a(r_a)}{1-\lambda}\\
&= \mathit{Revenue}(\textsc{opt-ir}_{N(\epsilon)})+\epsilon.
\end{aligned}
$$

Thus, the result follows. ∎

Appendix C Proof of Theorem 2 (Sub-optimality of the SEQ mechanism)

We begin by presenting a useful result that will help us derive the sufficient condition for the sub-optimality of the SEQ mechanism.

LEMMA 1. *For any given* $\mathbf{r}$, $r_X \geqslant W(1;\mathbf{r}) \geqslant W(2;\mathbf{r}) \geqslant \ldots \geqslant W(N;\mathbf{r})$.

Proof: We will show through induction that $r_X \geqslant W(n;\mathbf{r}) \geqslant W(n+1;\mathbf{r})$, $\forall n \in \{1,2,\ldots,N\}$. Consider time slot N. We have $W(N;\mathbf{r}) = r_X\cdot p_{X,N} \leqslant r_X$ and $W(N;\mathbf{r}) \geqslant W(N+1;\mathbf{r})$. Thus, the claim holds for $n=N$. Suppose the claim holds true for any $k\leqslant N$. That is $r_X \geqslant W(k;\mathbf{r}) \geqslant W(k+1;\mathbf{r})$. Then

$$W(k-1;\mathbf{r}) = r_X\cdot p_{X,k-1} + (1-p_{X,k-1})\cdot W(k;\mathbf{r})$$

$$
\begin{aligned}
&\geqslant W(k;\mathbf{r})\cdot p_{X,k-1}+(1-p_{X,k-1})\cdot W(k;\mathbf{r})\\
&=W(k;\mathbf{r}).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
W(k-1;\mathbf{r}) &= r_X\cdot p_{X,k-1}+(1-p_{X,k-1})\cdot W(k;\mathbf{r})\\
&\leqslant r_X\cdot p_{X,k-1}+(1-p_{X,k-1})\cdot r_X\\
&= r_X.
\end{aligned}
$$

Thus, $r_X \geqslant W(k-1;\mathbf{r}) \geqslant W(k;\mathbf{r})$ and the result follows. ■

Next, we use this result to prove Theorem 2.

Since $r_X p_{X,1} > r_Y p_{Y,1}$, it follows that $r_X p_{X,1}\delta^{n-1} > r_Y p_{Y,1}\delta^{n-1}$ and thus $r_X p_{X,n} > r_Y p_{Y,n}$ $\forall n \in \{1,2,\ldots,N\}$. Therefore, the SEQ mechanism allocates ad X in all the time slots. Furthermore, it is straightforward to see that the revenue to the ad-exchange from the SEQ mechanism is $W(1;\mathbf{r})$.

Recall that the OPT-IR mechanism is obtained by solving the following dynamic program:

$$
\begin{aligned}
\hat{R}(n;\mathbf{r}) &= \max_{a\in\mathcal{A}}\left\{r_a p_{a,n}+\lambda(1-p_{a,n})\hat{R}(n+1;\mathbf{r})\right\};\ 1\leqslant n\leqslant N,\ \hat{R}(N+1;\mathbf{r}):=0,\\
\pi_n^{\text{OPT-IR}}(\mathbf{r}) &= \arg\max_{a\in\mathcal{A}}\left\{r_a p_{a,n}+\lambda(1-p_{a,n})\hat{R}(n+1;\mathbf{r})\right\};\ 1\leqslant n\leqslant N.
\end{aligned}
$$

We now show that for $1\leqslant n\leqslant N$, the OPT-IR mechanism allocates ad Y in the first $K-1$ time slots and ad X in the remainder of the time slots. That is,

$$
\pi_n^{\text{OPT-IR}}(\mathbf{r}) = \begin{cases} X & \text{if } n\geqslant K\\ Y & \text{if } n<K.\end{cases}
$$

First, we will show through induction that for $K\leqslant n\leqslant N$,

- $W(n;\mathbf{r}) = \hat{R}(n;\mathbf{r})$,
- $\pi_n^{\text{OPT-IR}}(\mathbf{r}) = X$.

Consider time slot N.

$$
\begin{aligned}
&\hat{R}(N;\mathbf{r}) = \max_{a\in\mathcal{A}}\{\psi(r_a)p_{a,N}\} = r_X p_{X,N} = W(N;\mathbf{r})\\
\Longrightarrow\ &\pi_N^{\text{OPT-IR}}(\mathbf{r}) = X.
\end{aligned}
$$

Suppose the claim holds true for any k, $K<k\leqslant N$. Consider period $k-1$. Since $k>K$, it follows that $\Delta(\mathbf{r})\geqslant W(k;\mathbf{r})$. Thus,

$$
\frac{r_X p_{X,1}-r_Y p_{Y,1}}{p_{X,1}-p_{Y,1}} \geqslant W(k;\mathbf{r})
$$

$$
\begin{aligned}
&\implies r_X p_{X,1} - r_Y p_{Y,1} \geqslant (p_{X,1} - p_{Y,1})\, W(k;\mathbf{r}) \\
&\implies r_X p_{X,1} - p_{X,1} W(n;\mathbf{r}) \geqslant r_Y p_{Y,1} - p_{Y,1} W(k;\mathbf{r}) \\
&\implies r_X \left(p_{X,1}\delta^{k-2}\right) - \left(p_{X,1}\delta^{k-2}\right) W(k;\mathbf{r}) \geqslant r_Y \left(p_{Y,1}\delta^{k-2}\right) - \left(p_{Y,1}\delta^{k-2}\right) W(k;\mathbf{r}) \\
&\implies r_X \left(p_{X,1}\delta^{k-2}\right) + \left(1 - p_{X,1}\delta^{k-2}\right) W(k;\mathbf{r}) \geqslant r_Y \left(p_{Y,1}\delta^{k-2}\right) + \left(1 - p_{Y,1}\delta^{k-2}\right) W(k;\mathbf{r}) \\
&\implies r_X \left(p_{X,k-1}\right) + \left(1 - p_{X,k-1}\right) W(k;\mathbf{r}) \geqslant r_Y \left(p_{Y,k-1}\right) + \left(1 - p_{Y,k-1}\right) W(k;\mathbf{r}) \\
&\implies r_X \left(p_{X,k-1}\right) + \left(1 - p_{X,k-1}\right) \hat{R}(k;\mathbf{r}) \geqslant r_Y \left(p_{Y,k-1}\right) + \left(1 - p_{Y,k-1}\right) \hat{R}(k;\mathbf{r}) \\
&\implies \hat{R}(k-1;\mathbf{r}) = r_X \left(p_{X,k-1}\right) + \left(1 - p_{X,k-1}\right) \hat{R}(k;\mathbf{r}) \\
&\qquad = W(k-1;\mathbf{r}).
\end{aligned}
$$

Consequently, $\pi_n^{\text{OPT-IR}}(\mathbf{r}) = X$ for all $K < k \leqslant N$.

Now consider time slot K. By definition, $W(K;\mathbf{r}) > \Delta(\mathbf{r})$. Thus,

$$
\begin{aligned}
&\frac{r_X p_{X,1} - r_Y p_{Y,1}}{p_{X,1} - p_{Y,1}} < W(K;\mathbf{r}) \\
&\implies r_X p_{X,1} - r_Y p_{Y,1} < (p_{X,1} - p_{Y,1})\, W(K;\mathbf{r}) \\
&\implies r_X p_{X,1} - p_{X,1} \lambda W(K;\mathbf{r}) < r_Y p_{Y,1} - p_{Y,1} W(K;\mathbf{r}) \\
&\implies r_X \left(p_{X,1}\delta^{K-2}\right) - \left(p_{X,1}\delta^{K-2}\right) W(K;\mathbf{r}) < r_Y \left(p_{Y,1}\delta^{K-2}\right) - \left(p_{Y,1}\delta^{K-2}\right) W(K;\mathbf{r}) \\
&\implies r_X \left(p_{X,1}\delta^{K-2}\right) + \left(1 - p_{X,1}\delta^{K-2}\right) W(K;\mathbf{r}) < r_Y \left(p_{Y,1}\delta^{K-2}\right) + \left(1 - p_{Y,1}\delta^{K-2}\right) W(K;\mathbf{r}) \\
&\implies r_X \left(p_{X,K-1}\right) + \left(1 - p_{X,K-1}\right) W(K;\mathbf{r}) < r_Y \left(p_{Y,K-1}\right) + \left(1 - p_{Y,K-1}\right) W(K;\mathbf{r}) \\
&\implies r_X \left(p_{X,K-1}\right) + \left(1 - p_{X,K-1}\right) \hat{R}(K;\mathbf{r}) < r_Y \left(p_{Y,K-1}\right) + \left(1 - p_{Y,K-1}\right) \hat{R}(K;\mathbf{r}) \\
&\implies \hat{R}(K-1;\mathbf{r}) = r_Y \left(p_{Y,K-1}\right) + \left(1 - p_{Y,K-1}\right) \hat{R}(K;\mathbf{r}).
\end{aligned}
$$

Consequently, $\pi_{K-1}^{\text{OPT-IR}}(\mathbf{r}) = Y$. Furthermore, $\hat{R}(K-1;\mathbf{r}) > W(K-1;\mathbf{r})$. It is easy to verify that $\hat{R}(n;\mathbf{r}) \geqslant \hat{R}(K-1;\mathbf{r}) > \Delta(\mathbf{r})$ and $\hat{R}(n;\mathbf{r}) > W(n;\mathbf{r}) \quad \forall n \in \{1, 2, \ldots, K-1\}$. Thus, it follows that $\pi_n^{\text{OPT-IR}}(\mathbf{r}) = Y\ \forall n \in \{1, 2, \ldots, K-1\}$ and therefore $\pi_n^{\text{OPT-IR}}(\mathbf{r}) = \begin{cases} X & \text{if } n \geqslant K \\ Y & \text{if } n < K \end{cases}$. Since $\hat{R}(1;\mathbf{r})$ is the revenue from the OPT-IR mechanism and $\hat{R}(1;\mathbf{r}) > W(1;\mathbf{r})$, the result follows. ■

Appendix D Proof of Theorem 3 (Optimality of the SEQ mechanism)

Proof: We use Theorem 1 to derive the OPT-IR mechanism through backward induction. For any $\mathbf{r}$, define

$$V(n;\mathbf{r}) := r_1 \cdot p_{1,n} + (1 - p_{1,n}) \cdot \lambda V(n+1;\mathbf{r});\ 1 \leqslant n \leqslant N,\ V(N+1;\mathbf{r}) = 0.$$

That is, $V(n;\mathbf{r})$ denotes the revenue obtained from period n onwards by displaying ad 1 in each time slot. Since $r_1 \geqslant \max_{a\in\mathcal{A}}\ \{r_a\}$ and $p_{1,n} \geqslant \max_{a\in\mathcal{A}}\ \{p_{a,n}\}\ \forall n \in \{1,2,\ldots,N\}$, it follows that

$$r_1 \cdot p_{1,n} \geqslant \max_{a\in\mathcal{A}}\ \{r_a \cdot p_{a,n}\}\ \forall n \in \{1,2,\ldots,N\}\,.$$

Thus, the optimal SEQ mechanism selects ad 1 to display in all the time slots. That is, $\pi_n^{\text{SEQ}}(\mathbf{r}) = 1\ \forall n \in \{1,2,\ldots,N\}$. Further, it is straightforward to see that the revenue from the SEQ mechanism is $V(1;\mathbf{r})$. Next, we show that $\pi_n^{\text{OPT-IR}}(\mathbf{r}) = 1\ \forall n \in \{1,2,\ldots,N\}$.

Recall that the OPT-IR mechanism is obtained by solving the following dynamic program:

$$\hat{R}(n;\mathbf{r}) = \max_{a\in\mathcal{A}} \left\{ r_a p_{a,n} + \lambda(1-p_{a,n})\hat{R}(n+1;\mathbf{r}) \right\};\ 1 \leqslant n \leqslant N,\ \hat{R}(N+1;\mathbf{r}) := 0.$$

We now prove the following claims for all $n \in \{1,2,\ldots,N\}$ through backward induction:

1. $r_1 \geqslant V(n;\mathbf{r})$.
2. $V(n;\mathbf{r}) = \hat{R}(n;\mathbf{r})$.
3. $\pi_n^{\text{OPT-IR}}(\mathbf{r}) = 1$.

Consider time slot N. We have,

$$\begin{aligned} V(N;\mathbf{r}) &= r_1 \cdot p_{1,N} \leqslant r_1 \text{ (since } 0 \leqslant p_{1,N} \leqslant 1),\\ V(N;\mathbf{r}) &= r_1 \cdot p_{1,N} = \max_{a\in\mathcal{A}}\ \{r_a \cdot p_{a,N}\} \text{ (since } r_1 \geqslant \max_{a\in\mathcal{A}}\ \{r_a\} \text{ and } p_{1,N} \geqslant \max_{a\in\mathcal{A}}\ \{p_{a,N}\})\\ &= \hat{R}(N;\mathbf{r}). \end{aligned}$$

Consequently, $\pi_N^{\text{OPT-IR}}(\mathbf{r}) = 1$. Thus, the claim holds for $n = N$. Suppose the claim holds for any $k, k \leqslant N$. That is,

1. $r_1 \geqslant V(k;\mathbf{r})$.
2. $V(k;\mathbf{r}) = \hat{R}(k;\mathbf{r})$.
3. $\pi_k^{\text{OPT-IR}}(\mathbf{r}) = 1$.

Since $r_1 \geqslant V(k;\mathbf{r})$, it follows that $r_1 \geqslant r_1 \cdot p_{1,k-1} + (1-p_{1,k-1}) \cdot \lambda V(k;\mathbf{r})$ as $0 \leqslant p_{a,n} \leqslant 1\ \forall n$ and $0 \leqslant \lambda \leqslant 1$. Thus, $r_1 \geqslant V(k-1;\mathbf{r})$.

Case 1: $\max_{a\in\mathcal{A}\setminus\{1\}}\ r_a \geqslant \lambda V(k;\mathbf{r})$.

$$\begin{aligned} r_1 \cdot p_{1,k-1} + (1-p_{1,k-1}) \cdot \lambda V(k;\mathbf{r}) &\geqslant \max_{a\in\mathcal{A}\setminus\{1\}}\ \{r_a \cdot p_{1,k-1} + (1-p_{1,k-1}) \cdot \lambda V(k;\mathbf{r})\}\\ &= \max_{a\in\mathcal{A}\setminus\{1\}}\ \{p_{1,k-1}\,(r_a - \lambda V(k;\mathbf{r})) + \lambda V(k;\mathbf{r})\}\\ &\geqslant \max_{a\in\mathcal{A}\setminus\{1\}}\ \{p_{a,k-1}\,(r_a - \lambda V(k;\mathbf{r})) + \lambda V(k;\mathbf{r})\}\\ &= \max_{a\in\mathcal{A}\setminus\{1\}}\ \{r_a \cdot p_{a,k-1} + (1-p_{a,k-1}) \cdot \lambda V(k;\mathbf{r})\}. \end{aligned}$$

$$
\begin{aligned}
\text{Therefore,}\;\; \hat{R}(k-1;\mathbf{r}) &= \max_{a\in\mathcal{A}}\left\{r_a p_{a,k-1} + \lambda(1-p_{a,k-1})\hat{R}(k;\mathbf{r})\right\}\\
&= r_1\cdot p_{1,k-1} + (1-p_{1,k-1})\cdot\lambda V(k;\mathbf{r})\\
&= \hat{V}(k-1;\mathbf{r}),\;\text{and}\\
\pi^{\text{OPT-IR}}_{k-1}(\mathbf{r}) &= 1.
\end{aligned}
$$

Case 2: $\max_{a\in\mathcal{A}\setminus\{1\}}\; r_a \leqslant \lambda V(k;\mathbf{r})$.

$$
\begin{aligned}
r_1\cdot p_{1,k-1} + (1-p_{1,k-1})\cdot\lambda V(k;\mathbf{r}) &\geqslant \lambda V(k;\mathbf{r})\cdot p_{1,k-1} + (1-p_{1,k-1})\cdot\lambda V(k;\mathbf{r})\\
&= \lambda V(k;\mathbf{r})\\
&\geqslant \max_{a\in\mathcal{A}\setminus\{1\}}\;\{r_a\cdot p_{a,k-1} + (1-p_{a,k-1})\cdot\lambda V(k;\mathbf{r})\}.\\
\text{Therefore,}\;\; \hat{R}(k-1;\mathbf{r}) &= \max_{a\in\mathcal{A}}\left\{r_a p_{a,k-1} + \lambda(1-p_{a,k-1})\hat{R}(k;\mathbf{r})\right\}\\
&= r_1\cdot p_{1,k-1} + (1-p_{1,k-1})\cdot\lambda V(k;\mathbf{r})\\
&= \hat{V}(k-1;\mathbf{r}),\;\text{and}\\
\pi^{\text{OPT-IR}}_{k-1}(\mathbf{r}) &= 1.
\end{aligned}
$$

Thus, the claim holds true for all $n\in\{1,2,\ldots,N\}$. Since the revenue from the OPT-IR mechanism is $\hat{R}(1;\mathbf{r})$ and $\hat{R}(1;\mathbf{r}) = V(1;\mathbf{r})$, the result follows. ■

Appendix E Proof of Theorem 4

The mechanism

$$
\begin{aligned}
\mathbf{\Pi}^{\text{BASE}}(\mathbf{r}) &= \arg\max_{\boldsymbol{\mu}}\left\{\sum_{a=1}^{A}\psi_a(r_a)\Theta^{\boldsymbol{\mu}}_a(\mathbf{r})\right\},\;\text{s.t. (12)},\\
M^{\text{BASE}}_a(\mathbf{r}) &= r_a\Theta^{\text{BASE}}_a(\mathbf{r}) - \int_0^{r_a}\Theta^{\text{BASE}}_a(t_a,\mathbf{r}_{-a})\,dt_a\;\;\forall a\in\mathcal{A}.
\end{aligned}
$$

can be further simplified as follows. Using the equality constraints in (12), $\Theta^{\boldsymbol{\mu}}_a(\mathbf{r})$ can be written as:

$$
\begin{aligned}
\Theta^{\boldsymbol{\mu}}_a(\mathbf{r}) &= \sum_{n=1}^{N} p_{a,n}\mathbb{1}\{\pi^{\boldsymbol{\mu}}_n(\mathbf{r}) = a\}\,\lambda^{n-1}\prod_{t=1}^{n-1}(1-p_{\pi^{\boldsymbol{\mu}}_t(\mathbf{r}),t})\;\;\forall a\in\mathcal{A}\\
&= \mathbb{1}\{\pi^{\boldsymbol{\mu}}_1(\mathbf{r}) = a\}\sum_{n=1}^{N} p_{a,n}\lambda^{n-1}\prod_{t=1}^{n-1}(1-p_{a,t})\;\;\forall a\in\mathcal{A}.
\end{aligned}
\tag{22}
$$

Using (22), the allocation rule under the BASE mechanism, $\mathbf{\Pi}^{\text{BASE}}$, can be obtained by solving:

$$
\pi^{\text{BASE}}_n(\mathbf{r}) = \arg\max_{a\in\mathcal{A}^+}\;\left\{\psi_a(r_a)\sum_{n=1}^{N} p_{a,n}\lambda^{n-1}\prod_{t=1}^{n-1}(1-p_{a,t})\right\};n = 1,2,...,N. \tag{23}
$$

Let

$$y_a(\mathbf{r}_{-a}) := \inf\left\{t_a : \psi_a(t_a) \geqslant 0 \text{ and } \forall\ j \neq a, \psi_a(t_a)\Theta_a^{\text{BASE}}(t_a, \mathbf{r}_{-a}) \geqslant \psi_j(r_j)\Theta_j^{\text{BASE}}(t_a, \mathbf{r}_{-a})\right\}.$$

$y_a(\mathbf{r}_{-a})$ denotes the smallest bid that advertiser a has to make to win against the bid vector $\mathbf{r}_{-a}$. Using (23) and the definition of $y_a(\mathbf{r}_{-a})$, we have

$$\Theta_a^{\text{BASE}}(t_a, \mathbf{r}_{-a}) = \begin{cases} \sum_{n=1}^{N} p_{a,n}\lambda^{n-1} \prod_{t=1}^{n-1}(1-p_{a,t}) & \text{if } t_a \geqslant y_a(\mathbf{r}_{-a}), \\ 0 & \text{if } t_a < y_a(\mathbf{r}_{-a}). \end{cases}$$

Therefore,

$$\int_0^{r_a} \Theta_a^{\text{BASE}}(t_a, \mathbf{r}_{-a})dt_a = \begin{cases} (r_a - y_a(\mathbf{r}_{-a})) \sum_{n=1}^{N} p_{a,n}\lambda^{n-1} \prod_{t=1}^{n-1}(1-p_{a,t}) & \text{if } r_a \geqslant y_a(\mathbf{r}_{-a}), \\ 0 & \text{if } r_a < y_a(\mathbf{r}_{-a}). \end{cases}$$

Consequently, $\forall a \in \mathcal{A}$, we get:

$$M_a^{\text{BASE}}(\mathbf{r}) = \begin{cases} y_a(\mathbf{r}_{-a}) \cdot \sum_{n=1}^{N} p_{a,n}\lambda^{n-1} \prod_{t=1}^{n-1}(1-p_{a,t}) & \text{if } r_a \geqslant y_a(\mathbf{r}_{-a}), \\ 0 & \text{if } r_a < y_a(\mathbf{r}_{-a}). \end{cases}$$

Since $r_a \geqslant y_a(\mathbf{r}_{-a}) \iff \pi_1^{\text{BASE}} = a$, the payment rule can be succinctly written as:

$$M_a^{\text{BASE}}(\mathbf{r}) = \Theta_a^{\text{BASE}}(\mathbf{r}) \cdot y_a(\mathbf{r}_{-a})\ \forall a \in \mathcal{A}. \tag{24}$$

Thus, the mechanism $(\mathbf{\Pi}^{\text{BASE}}, \mathbf{M}^{\text{BASE}})$ specified by (23) and (24) is optimal for problem (P^{BASE}). ■

Appendix F Calculations for Win-Win Scenario (Discussed in Section 8.1)

For the BASE mechanism, the click-probabilities of ads X and Y across the two time slots are $2p + 0 = 2p$ and $p + (1-p)p \approx 2p$, respectively. Therefore, we have

$$\pi_1^{\text{BASE}}(\mathbf{r}) = \begin{cases} X & \text{if } v_X \geqslant v_Y, \\ Y & \text{if } v_X < v_Y. \end{cases}$$

The expected utilities of advertisers X and Y are simply the information rents they receive. Thus,

$$U_X^{\text{BASE}}(r_X) = \begin{cases} \frac{1-v_X}{2} \cdot 2p & \text{if } v_X \geqslant v_Y, \\ 0 & \text{if } v_X < v_Y. \end{cases}$$

$$U_Y^{\text{BASE}}(r_Y) = \begin{cases} \frac{1-v_Y}{2} \cdot 2p & \text{if } v_Y \geqslant v_X, \\ 0 & \text{if } v_X < v_Y. \end{cases}$$

$$\text{Thus, } \mathbb{E}_{r_X}\left[U_X^{\text{BASE}}(r_X)\right] = \int_{v_Y=0}^{1} \int_{v_X=y}^{1} p(1-x)dxdy = \frac{p}{6}.$$

Similarly, $\mathbb{E}_{r_Y}\left[U_Y^{\text{BASE}}(r_Y)\right] = \frac{p}{6}$. Let $Rev^{\text{BASE}}(\mathbf{r})$ denote the revenue to the ad exchange under the BASE mechanism. Thus,

$$Rev^{\text{BASE}}(\mathbf{r}) = \begin{cases} \frac{1+v_Y}{2} \cdot 2p & \text{if } \; v_X \geqslant v_Y, \\ \frac{1+v_X}{2} \cdot 2p & \text{if } \; v_X < v_Y. \end{cases}$$

$$\text{Thus,} \;\; \mathbb{E}_{\mathbf{r}}\left[Rev^{\text{BASE}}(\mathbf{r})\right] = \int_0^1 \left[\int_0^y p(1+x)dx + \int_y^1 p(1+y)dx \right] dy = \frac{4p}{3}.$$

Next, we evaluate the expected utilities of the advertisers under the OPT-IR mechanism. Since $p_{X,2} = 0$ and $p_{Y,2} = p > 0$, it is clear that the OPT-IR mechanism selects ad Y to display in slot 2. Thus, $\pi_2^{\text{OPT-IR}}(\mathbf{r}) = Y$. In slot 1,

$$\pi_1^{\text{OPT-IR}}(\mathbf{r}) = \begin{cases} X & \text{if } \; v_X \cdot 2p + (1-2p) \cdot v_Y \cdot p \geqslant v_Y \cdot p + (1-p) \cdot v_Y \cdot p, \\ Y & \text{if } \; v_X \cdot 2p + (1-2p) \cdot v_Y \cdot p < v_Y \cdot p + (1-p) \cdot v_Y \cdot p. \end{cases}$$

Since $p^2 \approx 0$, we get

$$\pi_1^{\text{OPT-IR}}(\mathbf{r}) = \begin{cases} X & \text{if } \; 2v_X \geqslant v_Y, \\ Y & \text{if } \; 2v_X < v_Y. \end{cases}$$

Consequently,

$$U_X^{\text{OPT-IR}}(r_X) = \begin{cases} \frac{1-v_X}{2} \cdot 2p & \text{if } \; 2v_X \geqslant v_Y, \\ 0 & \text{if } \; 2v_X < v_Y. \end{cases}$$

$$U_Y^{\text{OPT-IR}}(r_Y) = \begin{cases} \frac{1-v_Y}{2} \cdot p & \text{if } \; 2v_X \geqslant v_Y, \\ \frac{1-v_Y}{2} \cdot 2p & \text{if } \; 2v_X < v_Y. \end{cases}$$

$$\text{Thus,} \;\; \mathbb{E}_{r_X}\left[U_X^{\text{OPT-IR}}(r_X)\right] = \int_0^1 \int_{\frac{y}{2}}^1 p(1-x)dxdy = \frac{7p}{24},$$

$$\mathbb{E}_{r_Y}\left[U_Y^{\text{OPT-IR}}(r_Y)\right] = \int_0^1 \left[\int_0^{\frac{y}{2}} p(1-y)dx + \int_{\frac{y}{2}}^1 \frac{p(1-y)}{2}dx \right] dy = \frac{7p}{24},$$

$$\mathbb{E}_{\mathbf{r}}\left[Rev^{\text{OPT-IR}}(\mathbf{r})\right] = \int_0^1 \left[\int_0^{\frac{y}{2}} 2pydx + \int_{\frac{y}{2}}^1 (2px+py)dx \right] dy = \frac{19p}{12}.$$

■

Appendix G Calculations for Win-Lose Scenario (Discussed in Section 8.2)

Consider the BASE mechanism. For $0 \leqslant q < \frac{3-\sqrt{6}}{3}$, we have

$$\begin{aligned} & 2q - q^2 \leqslant \frac{1}{3} \\ \implies & \frac{3}{2}(2q-q^2) \leqslant \frac{1}{2} \\ \implies & \frac{3}{2}(2q-q^2) \leqslant v_X \;\; \forall v_X \;\; (\text{since } v_X \sim U(0.5,1)). \end{aligned}$$

Thus, the BASE mechanism always selects ad X for display. The expected BASE utilities of advertisers X and Y, respectively, are $\mathbb{E}_{r_X}\left[U_X^{\text{BASE}}(r_X)\right] = \int_{0.5}^{1}\left(\frac{1-x}{2}\right)\cdot 2dx = \frac{1}{8}$ and $\mathbb{E}_{r_Y}\left[U_Y^{\text{BASE}}(r_Y)\right] = 0$. The expected revenue to the ad exchange is $\mathbb{E}_{\mathbf{r}}\left[Rev^{\text{BASE}}(\mathbf{r})\right] = \frac{3}{4}$.

Now consider the OPT-IR mechanism. In slot 2, since $v_X \geqslant v_Y \cdot q$ for all values of $\mathbf{r}$, we have $\pi_2^{\text{OPT-IR}}(\mathbf{r}) = X$. Further, since $v_Y \geqslant v_X$ for all $\mathbf{r}$, we have

$$v_Y q + (1-q)v_X \geqslant v_X.$$

Therefore, $\pi_1^{\text{OPT-IR}}(\mathbf{r}) = Y \ \forall \mathbf{r}$. Thus, we have $\mathbb{E}_{r_X}\left[U_X^{\text{OPT-IR}}(r_X)\right] = (1-q)\frac{3}{4}$ and $\mathbb{E}_{r_Y}\left[U_Y^{\text{OPT-IR}}(r_Y)\right] = 0$. The expected revenue to the ad exchange under the OPT-IR mechanism is then equal to $\mathbb{E}\left[v_Y \cdot q + (1-p)v_X\right] = \frac{3(1+q)}{4}$. ■