

A hierarchical framework for ad inventory allocation in programmatic advertising markets



Juanjuan Li^{a,b,c}, Xiaochun Ni^{b,c}, Yong Yuan^{b,c,*}, Fei-Yue Wang^{b,c}

^a School of Automation, Beijing Institute of Technology, Beijing, China

^b The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

^c Qingdao Academy of Intelligent Industries, Qingdao, China

ARTICLE INFO

Keywords:

Programmatic advertising
Ad inventory
Real-time bidding
Private marketplace
Header bidding

ABSTRACT

Enabled by the big data driven user profiling and precision bidding techniques, online programmatic advertising markets have evolved from the traditional website-buying or ad-slot-buying models to a fine-grained and real-time trading model at the level of ad impressions (i.e., ad inventory). As a result, Web publishers are now facing a challenging decision of allocating the ad inventory across multiple advertising models, which has a direct and important influence on both their individual revenues, and the market-wide supply-demand balance. In this paper, we propose a novel hierarchical ad inventory allocation framework (AIAF), taking into consideration the possible scenarios of ad inventory allocation in programmatic advertising markets. AIAF explicitly captures the specific features of ad inventory allocation in each of three levels (i.e., channel level, market level and platform level), and also their influence-feedback effects. We present the general solution process for solving this model on the basis of its property analysis. An illustrative instantiation of our AIAF model is formulated to demonstrate its applications in supporting publishers' decision-making on the ad inventory allocation. We also conduct experiments based on empirical data so as to validate the model and analysis. Our research findings indicate that 1) our AIAF model outperforms other single-level and two-level allocation strategies; 2) the fine-grained optimization is superior to that of the coarse-grained level; 3) allocation decisions should be made on the basis of the comparative marginal revenue instead of the absolute marginal revenue.

1. Introduction

Programmatic advertising (PA), as a novel format of online precision marketing, has sparked a new wave of explosive growth in display advertising markets. According to *eMarketer*, in USA, the PA promotion spending reaches 32.56 billion dollars in 2017, taking up the 80% market share of the online display advertising; in UK, programmatically traded ads account for more than 75% of online display advertising spending by the end of 2017; and in China, the PA market scale is about 11.69 billion dollars in 2017, and will grow to 29.6 billion in 2019 with the average growth rate of about 35%. Enabled by the big data driven user profiling and precision bidding techniques (Busch, 2015), online programmatic advertising markets have evolved from the traditional website-buying or ad-slot-buying models, to a fine-grained and real-time trading model at the level of ad impressions (i.e., ad inventory). This evolution of sales model has been witnessed to be able to facilitate precise matching between advertisements and target audiences in a

real-time fashion (Dawson et al., 2016), as well as effective allocation of the limited ad resources, thus leading to the improved performance in market promotions (Yuan et al., 2014).

The increasing prevalence of PA is widely recognized to attribute partially to the influx of millions of publishers, who play the critical role as ad impression suppliers in PA markets. In China, more than 90% online medias use PA to sell ad impressions according to the statistics of *iResearch*. Also, almost all the leading medias in China, e.g. Baidu, Alibaba, Tencent, Sina, etc. have marched into this field to adopt multiple ad models to sell the online advertising resources. One of the crucial decisions that arises for the publishers is ad inventory allocation, i.e. determining how to allocate the limited ad impressions to the demanding advertisers as to optimize the publishers' various objectives. The allocation decisions not only produce important influence on their revenue management, but also greatly affect the promotion performance, the supply-demand balance, and also the market structure in the PA market (Clerici and Perego, 2010; Li et al., 2018). As such, the ad

* Corresponding author at: The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

E-mail addresses: juanjuan.li@ia.ac.cn (J. Li), xiaochun.ni@ia.ac.cn (X. Ni), yong.yuan@ia.ac.cn (Y. Yuan), feiyue.wang@ia.ac.cn (F.-Y. Wang).

<https://doi.org/10.1016/j.elerap.2018.09.001>

Received 23 May 2018; Received in revised form 23 July 2018; Accepted 3 September 2018

Available online 05 September 2018

1567-4223/ © 2018 Elsevier B.V. All rights reserved.

inventory allocation is widely considered in the literature as a critical decision for publishers.

Generally speaking, publishers are confronted with three key challenges of uncertainty, scarcity and diversity in their decision-making tasks of the ad impression allocation. First, the market supply of ad impressions is uncertain due to the unpredictable user visits on the publishers' landing pages (Balseiro et al., 2014; Lai et al., 2017). Second, compared with advertisers' huge demand on ad display opportunities, current supply of ad impressions is still very limited according to the yearly report of *analysis*. Third, publishers typically sell the ad inventory by means of diversified models including online real-time bidding (RTB), private marketplace (PMP), header bidding (HB), and offline guaranteed contracts. Consequently, these challenges bring unprecedentedly higher complexity for publishers' decisions to allocate their ad impressions, which has a direct and important influence on both their individual revenues, and the market-wide supply-demand balance. Therefore, ad impressions allocation has been widely considered as one of the most critical decisions for publishers in PA markets (Muthukrishnan, 2009; Mostagir, 2010).

Besides the complex market environments, the ad impression allocation itself is a complex research issue for publishers. For instance, the limited ad inventory should be wisely allocated among multiple advertising models, which are remarkably different in trading mechanisms. Also, there exists strong coupling relationships between the ad inventory allocation and publishers' other strategies including pricing (Muthukrishnan, 2009), information disclosure (Cachon and Fisher, 2000; Li et al., 2017) and so on. Meanwhile, the ad impression allocation will significantly influence decisions of other downstream players in the PA ecosystems, such as the decision of demand side platforms (DSP) to match ad impressions with target audiences (Adikari and Dutta, 2015), due to the diversity in trading mechanisms (Stavrogianis et al., 2014), complexity of bidding dynamics (Adikari and Dutta, 2015) as well as the uncertainty in realized market equilibria (McAfee, 2011); Meanwhile, the advertisers' bidding strategies are also studied in the realistic context by constraints given by the ad inventory optimization (Fernandez-Tapia et al., 2016). To date, publishers have to deal with these challenges when allocating the ad resource using multiple models simultaneously, with the aim of dispersing risks and maximizing revenues. As such, there is a critical need for researchers to design a novel framework to help publishers make rational ad inventory allocation decisions in PA markets (Chen, 2013). This motivates our research.

In this paper, we first classify the four typical advertising models mentioned above into three levels, i.e., the channel level, the market level and the platform level, mainly based on their priority, granularity, and trading mechanism. We then propose a novel hierarchical ad inventory allocation framework (AIAF) model for studying publishers' three-level allocation strategies, which provides a basic architecture for analyzing allocation strategies in different levels and their influence-feedback effects with each other. The properties of our proposed AIAF are theoretically proved, and the general solution process is given for solving our AIAF model. An illustrative instantiation of AIAF is presented to give further explanation. Also, computational experiments based on real-world data are designed to validate our framework.

The rest of this paper is organized as follows. Section 2 reviews relevant literatures. Section 3 briefly states the ad inventory allocation problem, and then Section 4 formulates the AIAF model, analyzes its properties and presents the solution process. In Section 5, we present an instantiation of AIAF. Section 6 conducts computational experiments to validate the framework, and discusses the managerial insights of our research. Section 7 concludes.

2. Literature review

Ad inventory allocation is one of the most crucial tasks in almost all formats of online display advertising (Broder, 2008; Yang et al., 2010),

and can be categorized into the intensively studied resource allocation issue in Economics (Harris et al., 1982; Sudharshan, 1995; Krieger and Green, 2006). This issue mainly concerns the publisher's allocation decision to optimize efficiency or revenue while keep to pre-specified contracts (Agrawal and Wang, 2014; Feldman et al., 2009; Devanur and Hayes, 2009), which has been well discussed for the traditional display advertising and sponsored search auctions, but not for programmatic advertising. From the input perspective of allocation tasks, on one hand, the capacity of supplying ad impressions is typically constrained, which as a result leads to a relatively low flexibility and fault tolerance threshold in publishers' allocation decisions (Sabbaghi et al., 2014); From the output perspective, on the other hand, advertisers in PA markets only bid for their best-matched ad impressions, which limits the publishers' strategy space within the precisely targeted niche markets (Qin et al., 2017). These two facts thus impose significant challenges for publishers to make ad inventory allocation decisions with the purpose of revenue maximization.

In PA practice, publishers must leverage their revenues and the associated risks in different channels seeking for a trade-off solution, since offline channels typically have lower risks and long-run guaranteed revenues, while online channels on the contrary have higher risks but short-term higher revenues. More specifically, given a set of ad impressions and a general priori about demands arriving stochastically with associated properties, the publisher should decide a valid allocation whether and how to satisfy the demand to maximize total payoffs (Feldman et al., 2010). A popular method of controlling the ad inventory sales for revenue management is using the bid-price controls first introduced by Simpson (1989). This method control the ad inventory allocation through setting a threshold price for each advertiser, which can be interpreted as the opportunity cost of allocating one additional impression to the advertiser. A certainty equivalent control heuristic was used to discuss this trade-off problem by Roels & Fridgeirsdottir (Roels and Fridgeirsdottir, 2009) to show the necessity of adopting both channels to reach the global maximum revenue.

This trade-off problem is further compounded by the fact that the publisher is typically hard to access to all the information for ad inventory allocation. To address the scenario of unknown demand of online channel, Ghosh et al. (2009) considered the publisher as a bidder of a series of offline guaranteed contracts and bids on behalf of them; then the publisher dynamically accepts some offline requests and allocates corresponding ad inventory to them only when the auction is lost by online advertisers. They proved that randomized bidding is a useful compromise for the trade-off. Walsh et al. (2010) proposed an approach that can automatically partition ad inventory into abstract channels, and also proposed a new constraint generation algorithm for improving ad allocation. Vee et al. (2010) built a two-phase auction model to sample and compute a compact allocation plan, and assign ad impressions online to advertisers submitting contracts with overlapping targeting rules. They also provided a solution serving advertisers in an online manner that is provably nearly optimal. However, these works do not take consideration of the trade-off between the revenue from a spot channel adopting real-time bidding and the efficiency from the reservation-based channel. Balseiro et al. (2014) considered this trade-off problem under this setting and formulated it as a stochastic control problem to deal with combined optimization of multi-channel allocation. Their research showed that the jointly optimization over both channels brings in considerable advantages for the publishers.

Essentially, the main objective of publishers is to maximize revenue through appropriately managing their available ad inventory, and the pricing must reflect the considerations, thus the ad inventory allocation problem can also be casted as an ad inventory pricing problem (Radovanovic and Heavlin, 2012). The key issue of the ad inventory allocation is to predict advertisers' bids and determine proper sales prices of ad impressions selling through different channels, markets, and platforms (Wu et al., 2015; Yuan et al., 2013). However, it becomes tremendously hard for publishers to learn the distribution of bids and

sales prices, since proprietary algorithms of advertisers' auctions rely heavily on privately-owned target audience profile which might never be disclosed to publishers (Yuan et al., 2014). Also, for the reason that past decisions to show ads on a given inventory depend in part on the publisher's propensity to convert, the quality is hard to be determined by advertisers (Perlich et al., 2012), which further make challenges for the publisher to predict bids and make the price. When the online channel adopts several ways to sell ad inventory, the pricing method is combinatorial auctions (Cramton et al., 2006), and the pricing problem turns out to be an issue of finding a yield maximizing way to allocate the ad inventory. Fernandez-Tapia (2015) researched the optimization of RTB-based ad inventory buying under limited budget settings, and focused on the inventory pricing to obtain the optimal tactics on a straightforward way by solving a constrained optimization problem. Najafi Asadolahi and Fridgeirsdottir (2014) investigated the optimal pricing strategy for ad inventory when impressions and clicks are uncertain, and found that the general heuristics to convert between the Cost Per Click (CPC) and Cost Per Mille (CPM) pricing schemes may be misleading as it may cause a great amount of revenue loss for publishers. Chahuara et al. (2017) discussed the optimal setting of the reserve price for RTB auctions in a revenue maximization engine for the publisher. In particular, these previous works on ad inventory pricing are mainly focused on the online single-channel sales of ad inventory instead of the cross-channel allocation of ad inventory, thus are not enough to support the publishers' revenue management decisions.

To sum up, the current research efforts on the ad inventory management in PA markets is limited, and they focus on the allocation across channels, especially across online RTB markets and offline guaranteed contracts. The existing research efforts are supportive to the viewpoint that the joint optimization of cross-channel allocations can benefit the publishers in revenue maximization. However, they seldom take into consideration of the whole trading process of ad impressions in the PA market, as well as the other advertising models with more fine-grained levels (e.g. market level and platform level). That is, the state-of-the-art works of the ad inventory allocation strategies lack a systematic and structured consideration of the entire PA markets. Consequently, these research efforts are still far from enough to provide effective supports for publishers' decisions in practical scenarios. In view of this situation, we believe that it is necessary to propose a novel hierarchical framework incorporating the publishers' multi-level ad inventory allocation decisions. Moreover, the framework is designed to be capable of providing an open environment for diversified research proposals about ad inventory allocation models, strategies and algorithms.

3. Problem statement

Typically, an ad impression is triggered when a user visits the publisher's landing page in the PA market. Then, the publisher should make an immediate allocation decision to submit it to a certain advertising channel, market and platform. Considering the aforementioned four advertising models and the whole ad impression trading processes in the PA market, there exist three ad inventory allocation scenarios depicted in Fig. 1.

First, the publisher should allocate ad inventory across the offline and online channels. The offline channel is typically realized with the format of guaranteed contracts; while the online channel includes multiple advertising formats, which we will discuss below in the market level and the platform level. The demand and unit price (i.e., CPM) for ad impressions in the offline channel are predetermined as a consensus between publishers and advertisers through one-to-one negotiation; while the demand and sales price of the ad inventory in the online channel are determined by online auctions in a real-time fashion.

Then, for those ad impressions allocated to the online channel, the publisher continues to allocate them between the direct and indirect markets. Here, the direct market usually refers to the HB market, which

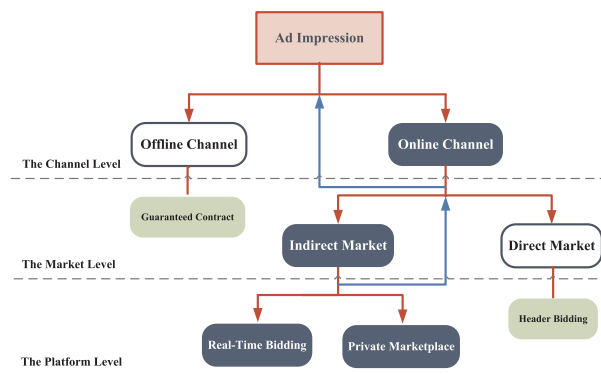


Fig. 1. The Three-level Ad Impression Allocation Scenarios in PA Markets.

enables the publisher to establish a priority sub-market to directly sell ad impressions to their allied advertisers. The indirect market trades ad impressions via Ad Networks or Ad Exchange (AdX), which serves as the intermediary to help the publisher deliver ad impressions to their precisely matched advertisers.

Finally, ad impressions allocated to the AdX should be further sold via either public AdX platforms or private AdX platforms. The public platform is built by the third-party agency, and adopts open auctions to sell ad impressions with diversified quality to generic advertisers. In contrast, the private platform is usually built by leading publishers and applies invitation-only private auctions to sell high-quality ad impressions to premium or VIP advertisers. Usually, the mainstream public and private platforms are RTB and PMP, respectively.

In view of the ad impression sales process in the PA market, we think it is of necessity to formulate a hierarchical AIAF model to deal with the three-level ad inventory allocation for publishers. First, according to above description of the ad impression sales practice, there exist the hierarchical relationships of the ad inventory sales through different ad models. Second, allocation decisions in three levels are inter-dependent, where upper levels create restrictions to the lower ones, and meanwhile lower levels produce feedbacks to the upper ones. Third, ad inventory allocation environments in each scenario are quite different, e.g., the channel-level allocation should consider the risk-revenue evaluation, the market-level allocation should consider the information disclosure, and the platform-level allocation is done under the principal-agent framework.

Accordingly, one of the contributions is designing a novel hierarchical three-level framework, corresponding to the above three ad inventory allocation scenarios. Our framework can serve as a generic model capable of dealing with the ad inventory allocation problems systematically. Then, we provide the solution process of AIAF on the basis of its property analysis. The other contribution is conducting the computational experiments to find the optimal ad inventory allocation strategies based on the real-world empirical data and validate the formulation of the AIAF model as well as the corresponding analysis.

4. Ad inventory allocation framework

In this section, we will propose a hierarchical AIAF model corresponding to the ad inventory allocation scenarios discussed in Section 3. Here, we consider a single publisher with the constrained supply of ad inventory and a numbers of advertisers without demand restrictions in the PA market. Table 1 describes the notations of this paper.

The **channel level** concerns the ad inventory allocation across offline and online channels; the **market level** focuses on the online ad inventory allocation between direct and indirect markets; and the **platform level** devotes to the ad inventory allocation among public and private platforms in indirect markets. Allocation strategies at these three levels are not separated but inter-dependent, forming a complete loop for the ad inventory allocation optimization where upper levels

Table 1
Summary of Notations.

Notations	Definitions
a	the total ad inventory
I	the number of channels
J	the number of markets
K	the number of platforms
a_i	the ad inventory allocated to the channel i
$a_{i,j}$	the ad inventory allocated to the market j in the channel i
$a_{i,j,k}$	the ad inventory allocated to the k th platform of the market j in the channel i
$p_{i,j,k}^x$	the sales price of the specific ad impression x in the k th platform of the j th market in the i th channel
$m_{i,j,k}^x$	the comparative marginal revenue of the publisher selling the ad impression x in the k th platform in the j th market of the i th channel
$c_{i,j,k}^x$	the commission rate of the ad impression x 's sales in the k th platform in the j th market of the i th channel
$e_{i,j,k}$	the AdXs' efforts to match ad impressions to targeted advertisers in the k th platform
$-k$	the competitive platforms of the AdX k
$m_{i,j}^x$	the comparative marginal revenue of the publisher selling the ad impression through the j 'th direct market
θ	the discounted factor
$-j'$	the competitive direct market of the direct market j'
$d_{i,j',k'}^x$	the unique information received by the advertiser k' about the ad impression x in the j' th direct market
$w_{i,j'}$	the unit variable cost of information disclosure in the j' th direct market
$b_{i,j',k'}^x$	the bid of the advertiser k' about the ad impression x in the j' th direct market
$D_{\hat{i}}$	the ad impression demand in the offline channel \hat{i}
$b_{\hat{i}}$	the unit price of ad impressions in the offline channel \hat{i}
$\beta_{\hat{i}}$	the penalty factor in the offline channel \hat{i}

create restrictions to the lower ones, and meanwhile lower levels produce feedbacks to the upper ones.

The channel-level allocation is the first decision faced by the publisher. Set the total ad inventory as a , and assume there are I alternative channels for the publisher. Then, the ad inventory allocation in the channel level can be given as:

$$a \rightarrow a_1, \dots, a_i, \dots, a_I, i \in \{1, 2, \dots, I\},$$

where a_i denotes the ad inventory allocated to the channel i . Generally, more ad inventories allocated to a certain channel will yield more revenues for the publisher from it; however, excessive allocation over a threshold may result in diseconomy for the reasons that 1) the offline guaranteed revenue has a fixed upper bond, and excessive supply creates no extra gainings; 2) too much supply for the online channel may result in comparatively low sales price or even sales failure when facing with budget-constrained advertisers and fierce competitions from other publishers.

The market-level decision is with regard to further allocation among segment markets in the online channel, which can be defined as:

$$a_i \rightarrow a_{i,1}, \dots, a_{i,j}, \dots, a_{i,J}, j \in \{1, 2, \dots, J\},$$

where $a_{i,j}$ denotes the ad inventory allocated to the market j in the channel i . The publisher's market-level ad inventory allocation decision is done under the constraint generated by the channel-level allocation result. Inversely, the market-level allocation will create effective feedbacks to the channel-level allocation.

In this level, the publisher determines the winning advertisers through online auctions. Trading mechanism, market structure and information disclosure are essentially different in direct and indirect markets. The direct market adopts the direct auctions between the publisher and advertisers, while the indirect market applies two-stage resale auction mechanism conducted by agents to match the publisher's supply and advertisers' demands. The publisher has more powerful control of trading process in the direct market than that in the indirect

market, since he/she could independently design the auction mechanism and information disclosure strategy in the direct market; while in the indirect market, these are greatly enslaved by various agents in the trading process. Therefore, the publisher should take essential differences between these two kinds of markets into consideration to decide the online allocation in the market level.

The platform-level allocation aims to decide which platforms should be connected to sell the ad inventory in the indirect market of the online channel, and it can be given as:

$$a_{i,j} \rightarrow a_{i,j,1}, \dots, a_{i,j,k}, \dots, a_{i,j,K}, k \in \{1, 2, \dots, K\},$$

Here, $a_{i,j,k}$ represents the ad inventory for the k th platform of the market j in the channel i . Similarly, the platform-level allocation is constrained by the market-level allocation result, and also creates feedbacks to the market-level allocation.

In both RTB and PMP, the publisher has a principal-agent relationship with AdX. However, due to the differences in constructor and service object of public and private platforms, the publisher has more powerful control in the private platform than the public platform. Correspondingly, the difference in efforts and incentives of agents will transmit to form different costs and revenues for the publisher to sell the ad inventory through public and private platforms.

4.1. Formulation

Inspired by the decentralized planning problems, we formulate AIAF as a hierarchical programming model taking three-level ad inventory allocation in the PA market into consideration. With the purpose of understanding the framework more intuitively and clearly, we formulate the model with the order from more fundamental fine-grained level to coarse-grained level, i.e., from platform level to market level and then to channel level.

Model 1 (Platform-level model) The ad impression allocation optimization in this level is subject to the constraint of ad impressions supply in the indirect market. The decision on the ad inventory allocation between RTB and PMP results in different principal-agent relationships.

Based on the analysis, we model the ad impression allocation problem in the platform level as:

$$\begin{aligned} z_{i,j}^{(1)} := \max_k & f_{i,j}^{(1)}(a_{i,j,k}) \\ \text{s. t.} & g_{i,j}^{(1)}(a_{i,j}, a_{i,j,k}) \leq 0, \\ & a_{i,j,k} \in A_1, A_1 \subset A_2. \end{aligned} \tag{1}$$

Here, $f_{i,j}^{(1)}$ is the payoff function in the platform level, which should be considered under the principal-agent relationships of publishers and AdXs. The supply of ad inventory in the platform level constrained by the market-level allocation result $a_{i,j}$ is defined by $g_{i,j}^{(1)}$.

Model 2 (Market-level model) The optimization of the ad inventory allocation in this level is subject to the constraint from allocation results of online channels. Selling ad inventory through direct and indirect markets leads to different auction mechanisms and information disclosure strategies in online auctions.

Based on the analysis, we model the ad impression allocation problem in the market level as:

$$\begin{aligned} z_i^{(2)} := \max_j & f_i^{(2)}(a_{i,j}, z_{i,j}^{(1)}) \\ \text{s. t.} & g_i^{(2)}(a_i, a_{i,j}) \leq 0, \\ & a_{i,j} \in A_2, A_2 \subset A_3. \end{aligned} \tag{2}$$

Here, $f_i^{(2)}$ is the payoff function of market-level allocation, and $g_i^{(2)}$ is the supply constraint of the ad inventory in this level. In this level, trading mechanisms and information structures should be introduced into the payoff function the distinguish the differences of these two types of markets.

Model 3 (Channel-level model) The allocation optimization problem in the channel level is studied under the total supply constraint of ad impressions generated by web-page visitors. Typically, the publisher is willing to sell limited ad impressions across both offline and online channels with the purpose of revenue maximization under a certain risk constraint or risk minimization given a certain revenue constraint. In this paper, we consider the former situation to model the ad impression allocation problem in the channel level as follows:

$$\begin{aligned}
 z^{(3)} := & \max_i f^{(3)}(a_i, z_i^{(2)}) \\
 \text{s. t. } & g^{(3)}(a_i, a) \leq 0, \\
 & h^{(3)}(a_i) \leq 0, \\
 & a_i \in A_3.
 \end{aligned} \tag{3}$$

The payoff function in the channel level is defined by $f^{(3)}$, and $g^{(3)}$ is the total supply constraint of the ad inventory. The risk constraint of allocating the ad inventory across channels is given as $h^{(3)}$, which is determined by the publisher's individual risk preference or risk tolerance.

The above three models deal with the ad inventory allocation problem in a hierarchical way, and cover almost all the possible allocation scenarios in the PA market. These models are not independent, and can be combined to form the integrated closed-loop structure of AIAF formulation. Currently, we establish the AIAF as a framework, and do not present the concrete model of it. Note that, the payoff functions $f_i^{(2)}$ and $f^{(3)}$ in the upper level could be formulated as the linear or non-linear function of the optimal payoff from the lower level, which greatly depends on the publisher's objective. Also, the constraint function of each level can also be linear or non-linear in AIAF.

On one hand, AIAF conducts joint studies of the three-level allocations by considering their influence-feedback effects and coupling relationships. The allocation results of upper levels constrain the allocations in lower levels, and inversely the allocation results in lower level models also produce feedbacks to that of upper levels. On the other hand, AIAF can deal with the ad inventory allocation in a specific level independently by neglecting the coupling relationships with other levels.

In addition, AIAF is capable of considering much more complicated situations of ad inventory allocation in the PA market. For example, it can capture the dynamic allocation at different temporal granularity for each level. After the publisher determines the ad inventory allocation across channel during a certain period (e.g., month, week, etc.), he/she aims to distribute the online ad inventory to multiple markets over d temporal slots (e.g., day) in the market level. Furthermore, with the outcome from the market level as constraints, the platform-level model not only deals with the dynamic ad inventory allocation $a_{i,j,k}(t)$, but also the real-time adjustment. The real-time adjustment is conducted through a control variable $\Delta_{i,j,k}(t)$ according to some performance indicators of ad impressions (e.g. bid numbers, sales price, etc.). Accordingly, the AIAF model can be extended as:

$$\begin{aligned}
 z_{i,j}^{(1)}(d) := & \max_k \sum_t f_{i,j}^{(1)}[a_{i,j,k}(t)] \\
 \text{s. t. } & g_{i,j}^{(1)}[a_{i,j}, a_{i,j,k}(t)] \leq 0, \\
 & a_{i,j,k}(t+1) = a_{i,j,k}(t) + \Delta_{i,j,k}(t), \\
 & a_{i,j,k} \in A_1, A_1 \subset A_2.
 \end{aligned} \tag{4}$$

$$\begin{aligned}
 z_j^{(2)} := & \max_d \sum_d f_j^{(2)}[a_{i,j}(d), z_{i,j}^{(1)}(d)] \\
 \text{s. t. } & g_j^{(2)}(a_i, a_{i,j}(d)) \leq 0, \\
 & a_{i,j} \in A_2, A_2 \subset A_3.
 \end{aligned} \tag{5}$$

$$\begin{aligned}
 z^{(3)} := & \max_i f^{(3)}[a_i, z_i^{(2)}] \\
 \text{s. t. } & g^{(3)}[a_i(s), a] \leq 0, \\
 & h^{(3)}(a_i) \leq 0, \\
 & a_i \in A_3.
 \end{aligned} \tag{6}$$

4.2. Solution

In what follows, we will conduct the theoretical analysis of these three models, and try to find useful properties of AIAF to serve for finding the solution of AIAF.

For the following analysis, first we make transformation of payoff functions of above three models. Let $v_{i,j}^{(1)} = -f_{i,j}^{(1)}$, and then solving original model (1) is equivalent to solving the programming $\tilde{z}^{(1)}$ with the objective to minimize $v_{i,j}$ under constraints $g_{i,j}^{(1)}$. Similarly, we let $v_i^{(2)} = -f_i^{(2)}$, $v^{(3)} = -f^{(3)}$ to transform the programming $z^{(2)}$ in the market level and $z^{(3)}$ in the channel level into $\tilde{z}^{(2)}$, $\tilde{z}^{(3)}$, respectively.

Theorem 1. *If $v_{i,j}^{(1)}$, and $g_{i,j}^{(1)}$ are convex functions, the programming $\tilde{z}^{(1)}$ is convex.*

Proof. Let R be the constraint set, and we have

$$R = \{a_{i,j,k} | g_{i,j}^{(1)} \leq 0\}$$

Since $g_{i,j}^{(1)}$ is the convex function, it is easy to deduce that R is a convex set.

Define $X_R(t)$ as

$$X_R(x) = \begin{cases} 0, & \text{if } x \in R \\ \infty, & \text{else.} \end{cases}$$

We can verify that $X_R(x)$ is a convex function.

Define $\hat{v}_{i,j}^{(1)} = v_{i,j}^{(1)} + X_R(x)$. Because $v_{i,j}^{(1)}$ and $X_R(x)$ are convex functions, $\hat{v}_{i,j}^{(1)}$ is also a convex function. It can be proved that $\min_k \hat{v}_{i,j}^{(1)}$ is convex, where we have

$$\hat{v}_{i,j}^{(1)} = \begin{cases} v_{i,j}^{(1)}, & \text{if } a_{i,j,k} \in R \\ \infty, & \text{otherwise.} \end{cases}$$

Since $\min_k v_{i,j}^{(1)} = \min_k \hat{v}_{i,j}^{(1)}$ for $a_{i,j,k} \in R$, we can prove that $\min_k \hat{v}_{i,j}^{(1)}$ is convex. Therefore, $\tilde{z}^{(1)}$ is convex. \square

Theorem 2. *If $v_{i,j}^{(1)}$, $g_{i,j}^{(1)}$, $v_i^{(2)}$, and $g_i^{(2)}$ are all convex functions, and $v_i^{(2)}$ is nondecreasing on $\tilde{z}^{(1)}$, then the programming $\tilde{z}^{(2)}$ is convex.*

Proof. Define R as:

$$R = \{a_{i,j} | g_i^{(2)} \leq 0\}$$

Since $g_i^{(2)}$ is the convex function, we can get that R is a convex set. Similarly, we can prove that $X_R(x)$ is convex.

Let $\hat{v}_i^{(2)} = v_i^{(2)} + X_R(x)$. Because $v_i^{(2)}$ is convex function and nondecreasing on $\tilde{z}^{(1)}$, $\hat{v}_i^{(2)}$ is also a convex function. Therefore, we have $\min_j \hat{v}_i^{(2)}$ is convex, where we have

$$\hat{v}_i^{(2)} = \begin{cases} v_i^{(2)}, & \text{if } a_{i,j} \in R \\ \infty, & \text{otherwise.} \end{cases}$$

Since $\min_j v_i^{(2)} = \min_j \hat{v}_i^{(2)}$ for $a_{i,j} \in R$, we can prove that $\min_j \hat{v}_i^{(2)}$ is convex. Therefore, $\tilde{z}^{(2)}$ is convex. \square

Theorem 3. *If $v_{i,j}^{(1)}$, $g_{i,j}^{(1)}$, $v_i^{(2)}$, $g_i^{(2)}$, $v^{(3)}$, $g^{(3)}$, $h^{(3)}$ are all convex functions, and $v_i^{(2)}$ is nondecreasing on $\tilde{z}^{(1)}$, $v^{(3)}$ is nondecreasing on $\tilde{z}^{(2)}$, then the programming $\tilde{z}^{(3)}$ is convex.*

Proof. Similar as in Theorem 1 and Theorem 2, we can prove that $\min_i v^{(3)}$ is convex. Furthermore, we deduce the programming $\tilde{z}^{(3)}$ is convex. \square

Theorem 4. *Under the condition that $\tilde{z}^{(1)}$, $\tilde{z}^{(2)}$, $\tilde{z}^{(3)}$ are all convex programming, if $s^{(1)*}$, $s^{(2)*}$, $s^{(3)*}$ are the optimal strategy in each level respectively, $s^* = (s^{(1)*}, s^{(2)*}, s^{(3)*})$ is the optimal three-level allocation strategy and vice versa.*

Proof. First, we prove the sufficiency of this theorem. Since $s^{(1)*}$, $s^{(2)*}$, $s^{(3)*}$ are the optimal strategy for each level, we can get the minimal revenue of the platform level, market level and channel level

in the programming $\tilde{z}^{(1)}, \tilde{z}^{(2)}, \tilde{z}^{(3)}$ as $v_{ij}^{(1)*}, v_i^{(2)*}, v^{(3)*}$, respectively. Therefore, for $\forall s^{(1)} \neq s^{(1)*}$, we have $v_{ij}^{(1)*} < v_{ij}^{(1)}$; for $\forall s^{(2)} \neq s^{(2)*}$, we have $v_i^{(2)*} < v_i^{(2)}$; and for $\forall s^{(3)} \neq s^{(3)*}$, we have $v^{(3)*} < v^{(3)}$.

Here, we have $v^{(3)*} = v^{(3)}(v_i^{(2)*}) = v^{(3)}[v^{(2)}(v_{ij}^{(1)*})]$. Since $\tilde{z}^{(1)}, \tilde{z}^{(2)}, \tilde{z}^{(3)}$ are all convex programming, we can prove that v^* is the minimal payoff. Due to the equivalence of z and \tilde{z} , we can deduce that $f^{(3)*} = -v^{(3)*}$ is the maximal revenue for original three-level allocation models. Therefore, $s^* = (s^{(1)*}, s^{(2)*}, s^{(3)*})$ is the optimal allocation strategy for the AIAF model.

Then, we prove the necessity of this theorem. If $s^* = (s^{(1)*}, s^{(2)*}, s^{(3)*})$ is the optimal allocation strategy for the AIAF model, it must be the optimal strategy of the three-level model comprised of the programming $\tilde{z}^{(1)}, \tilde{z}^{(2)}$, and $\tilde{z}^{(3)}$. If $s^{(1)*}$ is not the optimal strategy for platform-level allocation, there must exist another strategy $s^{(1)} \neq s^{(1)*}$ to result in $v_{ij}^{(1)} < v_{ij}^{(1)*}$. Due to the coupling of three-level allocations, $s^{(1)}$ in the platform level must transmit to formulate a new optimal allocation strategy $s^{(2)} \neq s^{(2)*}$ with the minimal payoff $v_i^{(2)} < v_i^{(2)*}$ in the market level and $s^{(3)} \neq s^{(3)*}$ in the channel level with the minimal payoff $v^{(3)} < v^{(3)*}$. Here, the result that $v^{(3)} < v^{(3)*}$ is contradictory with the condition that $v^{(3)*}$ is the minimal payoff under s^* . Therefore, $s^{(1)*}, s^{(2)*}, s^{(3)*}$ are the optimal platform-level, market-level and channel-level allocation strategies, respectively. \square

Based on the above research, the solution process to find the optimal three-level allocation strategy for the AIAF formulation is depicted as follows.

First, given the total supply a_{ij} in the platform level, we get the optimal allocation strategy across RTB and PMP platforms as $s^{(1)\#}$ and also the maximal revenue $z^{(1)\#}$ through solving the model 1.

Second, substitute $z^{(1)\#}$ into the market-level model 2, where the total supply is a_i . Then, solve the model to get the optimal allocation strategy across AdX and HB markets as $s^{(2)\#}$ with the maximal revenue $z^{(2)\#}$.

Third, substitute $z^{(2)\#}$ into the channel-level model 3 with the total ad inventory a , solve the model to find the optimal allocation strategy across online and offline channels as $s^{(3)*}$ and also the maximal revenue $z^{(1)*}$.

Inversely, substitute $s^{(3)*}$ to the model 2 to constrain the market-level allocation, and obtain the optimal strategy $s^{(2)*}$, then further solve the model 1 to get the optimal allocation strategy in the platform level as $s^{(1)*}$.

Finally, we obtain the optimal three-level allocation strategy $s^* = (s^{(1)*}, s^{(2)*}, s^{(3)*})$, and the maximal revenue $z^{(1)*}$ for the publisher to sell a ad impressions in the PA market.

5. Instantiation

In this section, we formulate an illustrative instantiation to make the further demonstration of AIAF. In the PA market, only the successfully sold ad impressions can generate revenues for the publisher, no matter through offline negotiations or online auctions. That is, the opportunity cost of selling ad impressions in the PA market is zero, and owing but not selling ad impressions is not profitable for the publisher.

The publisher's payoff from the platform level can be computed by the sum of each ad impression's revenue acquired from K AdXs.

$$f_{ij}(1) = \sum_x \sum_k p_{ij,k}^x a_{ij,k}^x (1 - c_{ij,k})$$

We denote $p_{ij,k}^x$ as the individual sales price for the specific ad impression x paid by the auction-winning advertiser from the k th AdX in the j th market of the i th channel, and its sales price is $\sum_k p_{ij,k}^x a_{ij,k}^x$. Also, the sales price from each platform is greatly influenced by the AdX's efforts $e_{ij,k}$ to match the ad impression to more competitive targeted advertisers. Accordingly, we have $p_{ij,k}^x = p(e_{ij,k}, x)$. Here, $a_{ij,k}^x$ is a piecewise function, and we have

$$a_{ij,k}^x = \begin{cases} 1, & \text{if } m_{ij,k}^x > 0 \\ \zeta(1) \in (0, 1), & \text{if } m_{ij,k}^x = 0 \\ 0, & \text{otherwise.} \end{cases}$$

Here, $m_{ij,k}^x$ is defined as the comparative marginal revenue of the publisher selling the ad impression through the k th AdX, and it is calculated by

$$m_{ij,k}^x = p_{ij,k}^x - c_{ij,k} - \max[p_{ij,-k}^x - c_{ij,-k}]$$

where $-k$ denotes the competitive platforms. We have $a_{ij,k} = \sum_x a_{ij,k}^x$, and $a_{ij,k}$ is the number of ad impressions allocated to the k th AdX in the platform level.

Generally, how many ad impressions are allocated to a certain platform is greatly influenced by the sales price achieved from it, which is concerned with the AdX's effort. The AdX k does not simply accept ad impressions, but compete for them and hope to win over more commission payment. The commission rate for the ad impression sales is $c_{ij,k}^x$, and usually the commission rate in an AdX is identical for all ad impressions, thus we have $c_{ij,k}^x = c_{ij,k}$.

Also, the AdX k gets payoff from serving as the agent of the publisher, which is

$$f_{ij,k}(2) = \sum_x p_{ij,k}^x a_{ij,k}^x c_{ij,k} - \epsilon(a_{ij,k}, e_k)$$

where, $\epsilon(a_{ij,k}, e_{ij,k})$ represents the effort cost of the AdX k for helping the publisher sell $a_{ij,k}$ ad impressions. It is well-known that the publisher has more powerful control in the private AdX, therefore, they would like to access to premium advertisers with the high-quality ad impressions through the private AdX, while connect with generic advertisers through the public AdX; therefore, the sales prices and commissions of selling ad impressions in PMP platforms are much higher than that in RTB platforms. Correspondingly, the effort cost of the private AdX is generally higher than that of the public AdX.

Based on the above analysis, the model (1) can be substantiated as a multi-objective optimization problem under the principal-agent framework:

$$z_{ij}^{(1)} = \left\{ \begin{array}{l} \max f_{ij}(1) \\ \max f_{ij,k}(2) \end{array} \middle| a_{ij,k} \in G \right\} \quad (7)$$

where, the feasible solution set G is determined by the constraints in the platform level, which is

$$G = \left\{ a_{ij,k} \middle| \begin{array}{l} \sum_k \sum_x a_{ij,k}^x - a_{ij} \leq 0, \\ c_{ij,k} \geq 0, p_{ij,k}^x \geq 0, \\ a_{ij} \geq 0. \end{array} \right\}$$

In the market level, we view the K platforms in the j th indirect market as a whole to substantiate the market-level model (2) as follows:

$$z_i^{(2)} := \max \sum_j z_{ij}^{(1)} + \sum_{j'} z'_{ij'} - \theta \sum_{j'} W_{ij'} [1 - \frac{z'_{ij'}}{z'_{ij'}(a_i)}] \\ \text{s. t. } \sum_j a_{ij} + \sum_{j'} a_{ij'} \leq a_i, \\ a_i \geq 0, 0 \leq \theta \leq 1. \quad (8)$$

Here, we have revenue gained from the direct market calculated by $\sum_{j'} z'_{ij'}$ through allocating $\sum_{j'} a_{ij'}$ ad impressions to it, and revenue from the indirect market by $\sum_j z_{ij}^{(1)}$ from selling $\sum_j a_{ij}$ ad impressions.

We define $1 - \frac{z'_{ij'}}{z'_{ij'}(a_i)}$ as the under-utilization of the direct market j' , and $W_{ij'}$ as the sunk fixed cost to establish it. Here, θ is the discounted factor, and $z'_{ij'}(a_i)$ represents the revenue gained from the direct market j' when all the ad impressions of the online channel i are allocated to these direct markets. Based on the above analysis, the publisher's total revenue should be subtracted by the under-utilization loss to formulate

the market-level payoff function.

Comparing with the ad impression sales through AdXs in the indirect market, the publisher has the capacity to apply the differentiated information disclosure strategies in the direct market, which will result in the asymmetric information structures for all advertisers. However, the publisher should pay extra information disclosure cost to do so. Therefore, we formulate the revenue in the HB market as

$$z'_{i,j'} = \sum_x p_{i,j'}^x a_{i,j'}^x - a_{i,j'} w_{i,j'}$$

where the number of ad impressions allocated to the j' th HB market is $a_{i,j'} = \sum_x a_{i,j'}^x$, and

$$a_{i,j'}^x = \begin{cases} 1, & \text{if } m_{i,j'}^x > 0 \\ \zeta(2) \in (0, 1), & \text{if } m_{i,j'}^x = 0 \\ 0, & \text{otherwise.} \end{cases}$$

Here, $m_{i,j'}^x$ is defined as the comparative marginal revenue of the publisher selling the ad impression through the j' th direct market, and it is calculated by

$$m_{i,j'}^x = p_{i,j'}^x - w_{i,j'} - \max[p_{i,-j'}^x - w_{i,-j'}, p_{i,j'}^x(1 - c_{i,j}^x)]$$

$$p_{i,j'}^x = \max_{k'} [b_{i,j',k'}^x | b_{i,j',k'}^x \leq \max_{k'} b_{i,j',k'}^x]$$

where $-j'$ denotes the competitive direct market.

In the direct market, the asymmetric information structure is operable for the publisher. In view of this, we define the unique information received by the advertiser k' about the ad impression x in the j' th direct market as $d_{i,j',k'}^x$. The advertiser then determines his/her bid $b_{i,j',k'}^x = b(d_{i,j',k'}^x)$ of the ad impression on the basis of the received information. Among all the submitted bids, the second highest one will be the final sales price $p_{i,j'}^x$ of the ad impression. Actually, the sales prices of ad impressions also partially reflect the differences between direct and indirect markets. Li et al. (2017) has proved that the asymmetric information structure can bring in much higher sales price for the publisher than the symmetric information structure, especially when the publisher sells both high-quality and low-quality ad impressions. The unit variable cost for information disclosure in the j' th HB market is $w_{i,j'}$, and different information disclosure strategy corresponds to different cost, that is, $w_{i,j'} = w(d_{i,j'})$.

The channel level concerns maximizing the total payoff gained from both online and offline channels. We use $z = \sum_i z_i^{(2)}$ to define the online revenue, and $\hat{z} = \sum_{\hat{i}} z_{\hat{i}}$ to define the offline revenue. Here, the offline revenue in the \hat{i} th offline channel can be computed by

$$z_{\hat{i}} = a_{\hat{i}} b_{\hat{i}} - (D_{\hat{i}} - a_{\hat{i}}) \beta_{\hat{i}}$$

where $D_{\hat{i}}$ and $b_{\hat{i}}$ represent the demand and unit price of ad impressions in the offline channel \hat{i} respectively, which are both predetermined by offline negotiation. The publisher needs to guarantee the delivery of ad impressions in the negotiated contract; otherwise the penalty will be incurred to compensate offline advertisers (Roels and Fridgeirsdotir, 2009; Vee et al., 2010). The penalty factor is defined by $\beta_{\hat{i}}$, which is also agreed in advance. The excessive supply more than $D_{\hat{i}}$ is not desirable, since it will not bring in more than negotiated revenues for the publisher.

The publisher's cross-channel allocation decision of a ad impressions is to choose the proper channel portfolio for revenue maximization under a risk constraint. The online risk is mainly from the online sales failure, while the offline risk is mainly from the offline default. Therefore, we formulate the risk constraint as follows:

$$E\{([z + \hat{z}] - [u(z) + u(\hat{z})])^2\} \leq U$$

where $u(z)$ and $u(\hat{z})$ are expected revenue in online and offline channels, respectively; and U is the risk tolerance of the publisher.

Based on the above analysis, the model (3) can be substantiated as:

$$\begin{aligned} z^{(3)} := \max \quad & z + \hat{z} \\ \text{s. t.} \quad & E\{([z + \hat{z}] - [u(z) + u(\hat{z})])^2\} \leq U, \\ & \sum_i a_i + \sum_{\hat{i}} a_{\hat{i}} \leq a, \\ & 0 \leq \sum_{\hat{i}} a_{\hat{i}} \leq D_{\hat{i}}, \\ & a_i \geq 0, a_{\hat{i}} \geq 0 \\ & D_{\hat{i}} \geq 0, b_{\hat{i}} \geq 0, \\ & a \geq 0, U \geq 0. \end{aligned} \tag{9}$$

Following the general process of solving the AIAF model in Section 4.2, we can also solve the instantiation and find the optimal ad-inventory allocation strategy for the publisher.

6. Experiments

6.1. Experimental data and scene

In this section, we design computational experiments to validate the above analysis. The basic data set is the field logs released by one of the leading RTB platforms in China (*iPinyou.com.cn*), which comprises records of more than 3 million ad impressions and 10 million bids. We conduct the data processing of the raw data set to support more reliable experimental process as well as results, which mainly includes identifying the redundant data, filtering the data with missing values and detecting the data with outlier values. All of these imperfect data are removed to ensure that each bid ID matches one ad impression and all data are with perfect information.

From the analysis of real-world winning bid logs in *iPinyou* platform, we find it fits with the log normal distribution, which also has been verified by the research in Cui et al. (2011) and Li and Guan (2014). In accordance with the market practice stated in the white book released by an authoritative Chinese advertising research institution (*RTBChina.com*) that the premium ad impressions in HB markets will win the highest bids with the smallest variance, and the higher-quality ad impressions in PMP platforms will win the higher bids with the smaller variance than that from the general-quality ad impressions sold in RTB platforms, we fit the distribution functions of winning bids in PMP platforms and HB markets in Fig. 2, which will be used to generate data sets for the following experiments.

We consider a publisher supplies about 100 thousands of ad impressions in the PA market. The allocation experiments are conducted in the scenario that he/she decides to allocate ad impressions across one guaranteed contract, one HB market, one RTB platform and one PMP

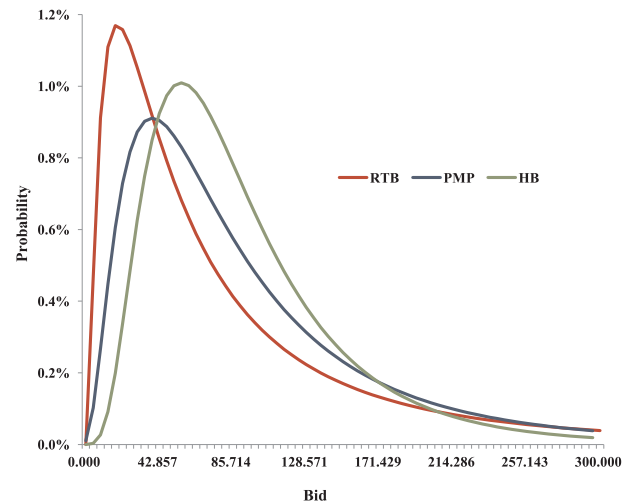


Fig. 2. Winning Bids (Unit: CNY) in RTB, PMP, and HB Fit Log-Normal Distribution.

platform. In detail, the publisher first determines the allocation across the offline guaranteed contract and the online channel, and the pre-determined offline demand is 20 thousands of ad impressions at CPM of CNY (Chinese Yuan) 50. Then, as for these ad impressions allocated to the online channel, he/she continues to allocate between the HB market and the indirect market, where ad impressions sold through HB has the higher information cost. Finally, he/she allocates ad impressions in the indirect market to the RTB platform and the PMP platform, where we simply consider the commission rates are determined by the AdXs, and the PMP platform has higher commission rates than the RTB platform does.

To ensure the reliability and validity of our experimental analysis, all experiments are conducted for 1000 times independently.

6.2. Three-level optimal allocation strategy

Under the experimental data set, given all parameters, we can figure out the three-level optimal allocation strategy for the publisher following the solution process described in Section 4.2. Under the optimal three-level allocation strategy, the maximal total revenue is CNY 12.051 million.

In the platform level, the optimal ad impressions allocated to PMP is 28,594, and to RTB is 22,132, and the publisher gets CNY 3.629 million from PMP and CNY 3.087 million from RTB. The results show that the publisher obtains more revenue through allocating more ad inventory to PMP.

For the optimal allocation in the platform level, we find an interesting phenomenon: the average unit revenue in PMP is lower than that in RTB; however, the average bid in PMP is much higher than that in RTB. This can be explained by the fact that a higher bid is accompanied by a higher cost in PMP. PMP guarantees the higher quality of ad impressions for advertisers and allows the publisher to have more control on the ad impression sales than RTB does. Therefore, although PMP requires more commissions transferred from the publisher, higher bids acquisition still makes it much more attractive than RTB in the platform level.

The platform-level allocation results tally with the practical situation in Chinese PA markets. According to the leading internet consulting organization in China (*iResearch*), PMP gradually surpasses RTB to be the dominant ad model in PA markets, and more and more publishers and advertisers access to PMP platforms.

In the market level, considering $\theta = 0$, the optimal strategy is to allocate 50,726 ad impressions to AdX and 39,780 to HB; and the maximal revenue from these two markets are CNY 6.716 million and CNY 4.912 million, respectively. Although HB does not get more ad inventory allocation than AdX, it wins more allocations compared with both RTB and PMP platforms.

Similar with the platform-level optimal allocation, the average unit revenue in HB is lower than that in AdX; however, the average bid from HB is higher than that from AdX. This phenomenon can also be explained by the fact that a higher bids is accompanied by a higher cost in HB. Since HB allows advertisers to access to publisher's premium ad inventory, meanwhile raises the charges of it for advertisers in terms of higher bids, and for the publisher in terms of growing information interaction costs. According to the experimental results, the balance of the cost and revenue results in more ad impressions sold through the indirect AdX market than the direct HB market.

In the real practice, the programmatic buying through direct markets is increasing while that through indirect markets is decreasing. In US, the direct buying takes up 56% share of the US programmatic advertising market till 2017 according to *eMarketer*; and in China, the direct buying gradually growing up and becoming more popular among advertisers according to *RTBChina*.

In the channel level, the optimal ad inventory allocated to the online channel is 90,506, and to the offline channel is 9494, and the maximal total revenue is CNY 12.051 million. The average unit revenue of all

online ad impressions is about CNY 128.211, which is much higher than the offline CPM. Therefore, the publisher prefers not to satisfy the offline contract but compensate the offline advertisers to strive for more ad impressions allocated to the online channel with the purpose of getting higher online revenues. However, completely ignoring of the offline contract is not feasible, since the online surplus is not enough to compensate such a large amount of offline penalties. Therefore, the publisher makes a balance between the online surplus and the offline penalty to decide the optimal allocation strategy in the channel level.

In practice, on one hand, publishers heavily rely on the online channel to boost short-term revenue, which is consistent with the gradually growing market share of PA models in the online digital display advertising market both in China and USA. On the other hand, publishers still attach much importance to offline advertisers in view of long-term revenue, therefore, it is irrational for them to completely abandon the offline channel.

The experimental results of the optimal allocation strategy also prove the validity of [Theorem 4](#) we propose in Section 4.2.

6.3. Comparative experiments

Furthermore, we design independent comparative experiments to validate the proposed AIAF model and the corresponding strategies. In what follows, we will compare our optimal AIAF strategy with the single-level and two-level optimal allocation strategies.

6.3.1. Single-level allocation strategy

First, we conduct experiments to compare the revenue maximization under the AIAF strategy with single-level allocation (SLA) strategies. In this paper, we do not consider the single-level optimization of the market level for the reason that very limited attentions have been paid for the market-level allocation by publishers in practice.

- The platform-level allocation strategy (the strategy SLA-P)

Under the strategy SLA-P, the publisher merely emphasizes the ad inventory allocation across RTB and PMP platforms for the revenue maximization. The publisher optimizes the platform-level allocation under the constraint determined by some commonly adopted upper-level allocation strategies, e.g., allocating all ad impressions to online channel (strategy AO) and fulfilling the offline contract and allocating the remaining to the online channel (strategy FO) in the channel level, allocating all to AdX (strategy AA) and the randomly average allocation between HB and AdX (strategy RA) in the market level.

[Table 2](#) depicts the optimization results of the strategy SLA-P under commonly adopted higher-level strategies mentioned above. From it, we can see that the revenue maximization under the strategy SLA-P is greatly influenced by the channel-level and market-level allocation strategies. The strategy SLA-P constrained by the strategy AO in the channel level and the strategy AA in the market level generates the highest optimal revenue, while it constrained by the strategy FO in the channel level and the strategy RA in the market level generates the lowest optimal revenue. Also, from the angle of the channel-level optimization, the strategy AO outperforms FO; and from the angle of the market-level optimization, the strategy AA outperforms RA. These results indicate that under the strategy SLA-P, more ad impressions are

Table 2
Total Revenue of Strategy SLA-P Under Typical Channel-Level And Market-Level Strategies (Unit: Million CNY)

Market-Level Strategy	Channel-Level Strategy	
	AO	FO
AA	9.478	8.744
RA	9.030	8.195

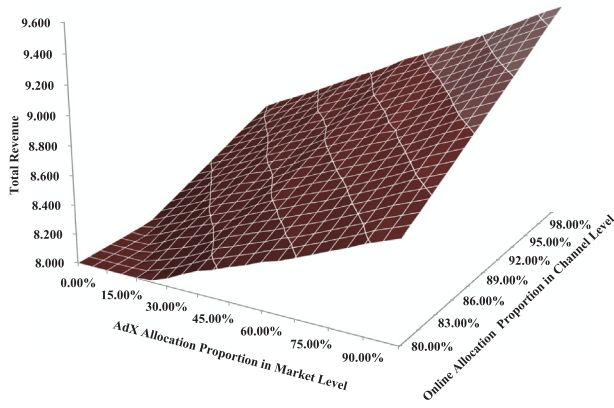


Fig. 3. Total Revenue Under Strategy SLA-P (Unit: Million CNY).

delivered to the AdXs, higher revenue will be achieved by the publisher.

To make in-depth studies on the strategy SLA-P, we further optimize total revenues of it under the generic channel-level and market-level allocation strategies, and the results are shown in Fig. 3. It is rational for us not to consider the excessive offline supply more than the pre-determined demand, therefore, the online allocation proportion will never be less than 80%. From Fig. 3, we can see that with the increasing allocations to the online channel in the channel level and the increasing allocations to AdX in the market level, the maximal total revenue under the strategy SLA-P increases, and the maxima of it is CNY 9.478 million achieved when all ad impressions are allocated to AdX, which is less than that under the optimal AIAF strategy.

- The channel-level allocation strategy (the strategy SLA-C)

Under the strategy SLA-C, the publisher focuses on the ad inventory allocation optimization across multiple channels and completely ignore the differences between these online channels. The publisher determines the cross-channel allocation relying on the comparison between the offline CPM and the online sales price. When the offline contract is not fulfilled, if the online payment is lower than the offline CPM, no matter whether it is the lowest one from HB, RTB and PMP, the ad impression will be allocated to the offline channel; if the online payment is higher than the offline CPM, no matter whether it is the highest one from HB, RTB and PMP, the publisher will decide to allocate it to the online channel only if the online payment exceeds the sum of the offline CPM plus the unit penalty. When the offline contract is fulfilled, the remaining impressions should be totally allocated to the online channel. The optimal SLA-C is to satisfy the offline demand first and then allocate the remaining to the online channel and the maximal revenue is CNY 3.986 million. From the result, we can find that the maximal revenue under the strategy SLA-C is far less than that under the AIAF strategy.

- Comparison of the strategy SLA-C and SLA-P

From comparison of the strategy SLA-C and SLA-P, we can find that 1): both strategies are inferior to the AIAF strategy in term of the revenue maximization, which means three-level allocation optimization is a better choice for the publisher than the single-level allocation optimization. 2) The strategy SLA-P surpasses SLA-C, which implies that optimization of the fine-grained level is much more profitable than that of the coarse-grained level. 3) The optimal allocation points under both strategies indicate that as much as possible ad inventory should be allocated to the level being optimized.

6.3.2. Two-level allocation strategy

Then, we conduct experiments to compare the revenue

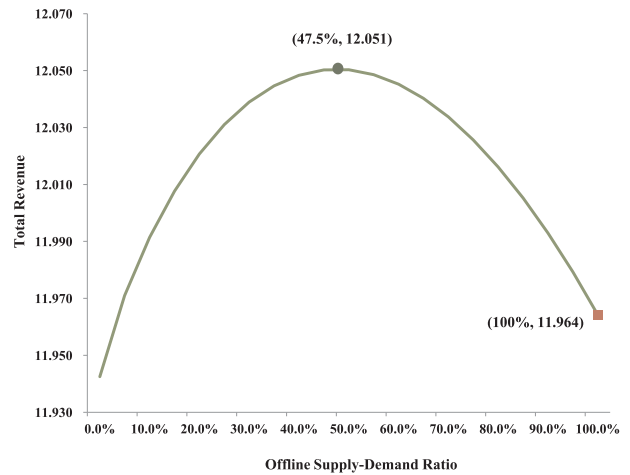


Fig. 4. Total Revenue Under Strategy TLA-MP (Unit: Million CNY).

maximization under the AIAF strategy with two-level allocation (TLA) strategies. In practice, joint optimization of channel level and platform level is usually not considered by publishers, therefore, we will neglect it and only consider the joint optimization of adjacent levels, i.e., market level and platform level, market level and platform level.

- Joint optimization of the market-level and platform-level allocation strategy (the strategy TLA-MP)

The strategy TLA-MP discusses joint optimization of the market-level and platform-level allocation regardless of the channel-level optimization.

The total revenue under the strategy TLA-MP is shown in Fig. 4, where the offline supply-demand ratio is defined by the offline supply dividing the offline demand. We can see that the maximal revenue is gained under the AIAF strategy for which the offline demand is partially satisfied in the proportion of 47.5%; while when the offline contract is completely fulfilled, the achieved revenue of CNY 11.964 million, which is much less than the maximal CNY 12.051 million.

From Fig. 4, we can see that the total revenue fits the trend of increasing quickly to a peak followed by a sharp decrease with the increasing offline allocation, which means under the soft constraint of the offline contract, both over-low and over-high offline allocation are not desirable. From the experiments, we can also conclude that the publisher can improve his/her optimal revenue through more effective offline negotiations to reduce the default penalty or agree on a more flexible offline demand.

Fig. 5 depicts the offline supply-demand ratio under the strategy TLA-MP with respect to the comparative marginal revenue, which is the sum of the offline marginal revenue and the online marginal loss (negative marginal revenue). Due to the fixed offline CPM and penalty factor, the offline marginal revenue is also fixed; but the online marginal loss increases with the growth of the offline allocation. When the offline marginal revenue can not compensate the online marginal loss, i.e. the comparative marginal revenue under Strategy TLA-MP reaches zero, the publisher should stop allocating more ad inventory to the offline channel. Therefore, the decision of the ad inventory allocation under the strategy TLA-MP mainly depends on the comparative marginal revenue.

- Joint optimization of the channel-level and market-level allocation strategy (the strategy TLA-CM)

The strategy TLA-CM discusses the two-level optimization of the channel-level and market-level ad inventory allocation while the platform-level optimization is not considered.

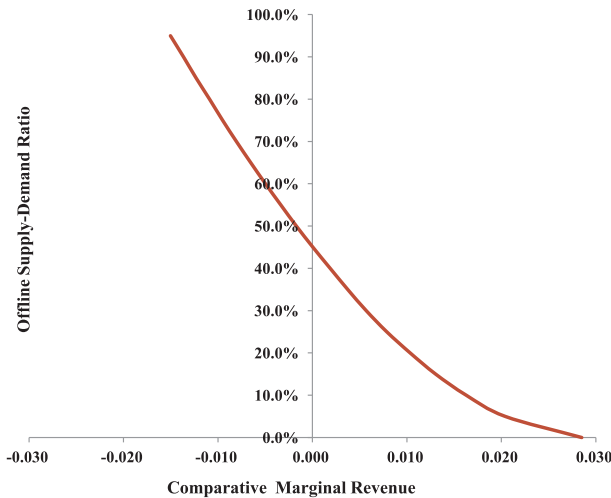


Fig. 5. Offline Supply-Demand Ratio with Comparative Marginal Revenue Under Strategy TLA-MP (Unit: Million CNY).

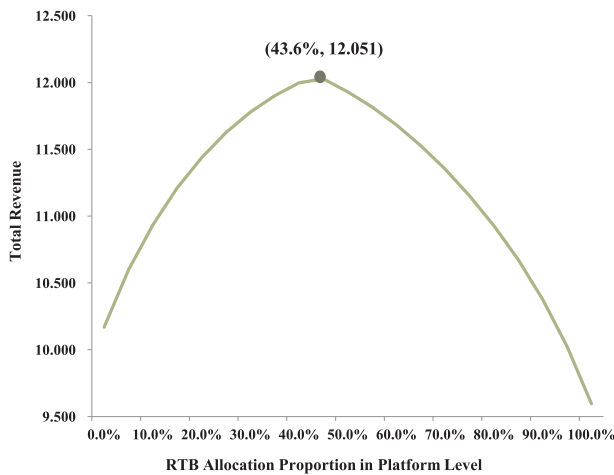


Fig. 6. Total Revenue Under Strategy TLA-CM (Unit: Million CNY).

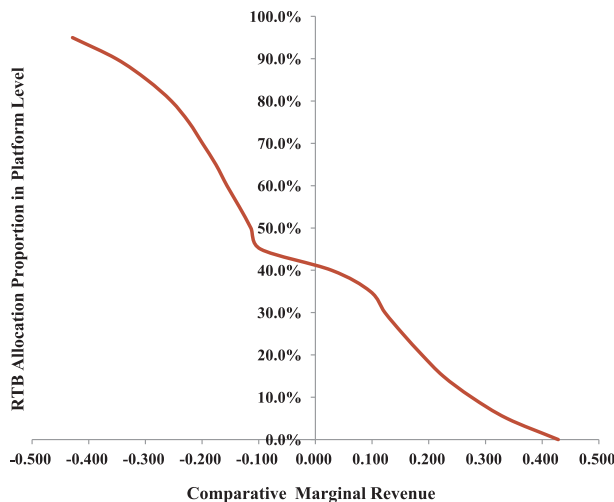


Fig. 7. RTB Allocation Proportion with Comparative Marginal Revenue Under Strategy TLA-CM (Unit: Million CNY).

Experimental results under the strategy TLA-CM are shown in Fig. 6 and 7. The optimal allocation under this strategy is equivalent to that under the optimal AIAF strategy, where 43.6% ad impressions in the

platform level are allocated to RTB, and the maximal revenue of CNY 12.051 million is achieved. Also, we can see that the allocation across two platforms outperforms that to the single platform in most cases, but it does not hold in all cases. When the allocation proportion of RTB reaches 92.2%, the total revenue is CNY 10.169 million, which is equal to that from allocating all ad inventory to PMP; and then more allocation to RTB will result in continual decreasing of the total revenue to the lowest CNY 9.596 million when all ad impressions are allocated to RTB. That is, when the RTB allocation proportion is in the interval (92.2%, 100%], the single-platform strategy of allocating all ad inventory to PMP is a better choice comparing to the cross-platform strategy.

From the joint analysis of the marginal revenue and the total revenue, we can see that when the comparative marginal revenue keeps positive, the publisher continues to allocate more ad inventory to RTB until it reaches zero, which is the optimal allocation point for the revenue maximization. When the allocation proportion of RTB reaches 43.6%, although the marginal revenue in RTB keeps positive and even gets increasing, it still can not compensate the increasing marginal loss in PMP, and the continual allocation to RTB will result in more loss of total revenue for the publisher. Therefore, the decision of the ad inventory allocation under the strategy TLA-CM is greatly influenced by the comparative marginal revenue.

As shown in Fig. 4 and 6, ad inventory allocation decisions abstracted from three-level AIAF models are concave, which conforms to the assumption of Theorem 1–4. As such, the analysis of the convex programming of the hierarchical AIAF models in Section 4.2 is verified.

• Comparison of the strategy TLA-MP and TLA-CM

Next, we further compare the strategy TLA-MP and TLA-CM. From Fig. 8, we can find that: 1) For these two strategies, the changing of total revenue shows a similar trend under different allocation proportions, but the strategy TLA-MP is much more stable than the strategy TLA-CM. 2) The lowest revenue under the strategy TLA-MP is CNY 11.942 million from allocating all ad inventory to the online channel; and the lowest revenue under the strategy TLA-CM is CNY 9.596 million from allocating all ad inventory in the platform level to RTB. 3) The average revenue under the strategy TLA-MP is CNY 12.018 million, and that under the strategy TLA-CM is CNY 11.178 million. Therefore, we can conclude that the strategy TLA-MP outperforms TLA-CM, since it can generate higher revenues for the publisher more stably.

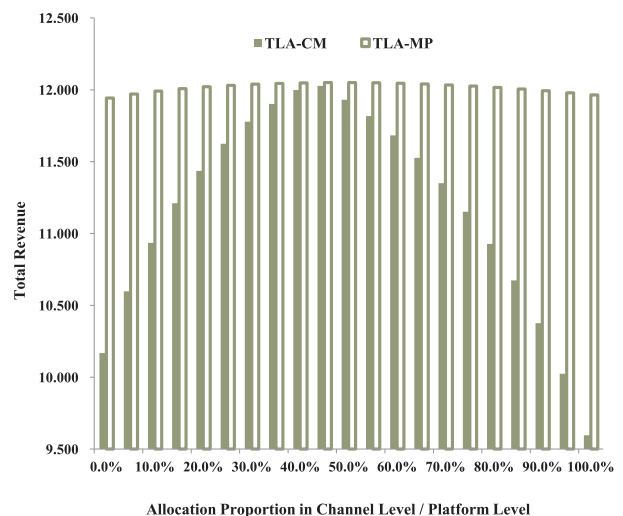


Fig. 8. Comparison of Total Revenue Under Strategy TLA-MP and Strategy TLA-CM (Unit: Million CNY).

6.4. Experimental discussions

From comparing the AIAF strategy with both the single-level allocation strategies and two-level allocation strategies, we can draw the following conclusions:

- 1) Comparing the experimental results in Section 6.2 and 6.3.1, we find that the optimal revenue of the single-level allocation strategies is far less than that of the three-level allocation strategy (i.e. our AIAF strategy). Also, comparing the experimental results in Section 6.2 and 6.3.2, we find that the upper bound of the optimal revenue achieved by the two-level allocation strategies is equal to the optimal revenue of the AIAF strategy. As such, we can conclude that our AIAF strategy outperforms both two-level allocation strategies and single-level allocation strategies. Accordingly, with the purpose of revenue maximization, the publisher should take the whole PA markets into consideration to conduct joint optimization of the ad inventory allocation in almost all scenarios.
- 2) Comparing the experimental results under the strategy SLA-P in Section 6.3.1 and the strategy TLA-MP in Section 6.3.2, we find that the revenues under the strategy TLA-MP are comprehensively higher than that under the strategy SLA-P. Similarly, comparisons of the experimental results under the strategy SLA-C and TLA-CM show that the strategy TLA-CM produces more revenues for the publisher than the strategy SLA-C. Moreover, the above analyses have highlighted the superiority of the AIAF strategy over the single-level allocation strategies. As such, for a specific level, joint optimization with other one or two levels outperforms the isolated optimization of itself significantly. That is, joint optimization of the multi-level allocation is a better choice than the single-level optimization for the publisher with limited ad inventory in PA markets.
- 3) For single-level allocation strategies, the strategy SLA-P outperforms SLA-C as discussed in Section 6.3.1; and regarding two-level allocation strategies, the strategy TLA-MP outperforms TLA-CM as discussed in Section 6.3.2, which means the fine-grained optimization of lower levels can generate higher revenues than the coarse-grained optimization of upper levels.
- 4) In case that more than 92.2% ad impressions in the market level are allocated to RTB, the platform-level optimization (i.e. under the strategy SLA-P) even results in more revenues than the joint optimization of the channel-level and market-level allocation (i.e. under the strategy TLA-CM), which indicates that sometimes the most fine-grained single-level optimization even take advantages over the coarser-grained joint optimization.
- 5) According to the analysis in Section 6.3, ad inventory allocation decisions should be made under the comparative marginal revenue from different channels, markets or platforms. If the marginal revenue of a specific channel, market or platform is high enough to compensate the marginal loss of the other, continuing the additional allocation to it is a good choice for the publisher. The optimal allocation strategy is realized when the marginal revenue and the marginal loss reaches zero-sum.

6.5. Managerial insights

Diversified ad models of PA markets suggest more choices for publishers to sell the limited ad inventory, but meanwhile add great complexities of allocation decisions. This research works on the three-level ad inventory allocation for publishers in PA markets and provides supportive managerial insights for their decisions.

First, the ad inventory allocation is a very significant decision for publishers, and slight changes of allocation strategies may result in significantly different revenues, especially for the accumulated effects of the three-level allocation.

Second, because of the hierarchical structure and strong couplings of three-level ad inventory sales, the ad inventory allocation can not be

simply considered as independent and isolated decisions in each level. Instead, publishers should take the influence-feedback effects of all three levels into consideration to make joint optimization with the purpose of revenue maximization. Our research has proved that the three-level allocation strategy always outperforms two-level allocation strategies, and similarly, two-level allocation strategies are better than the single-level allocation strategies.

Third, it is the usual practice for publishers and the common research interest of academia to put more efforts on more coarse-grained ad inventory allocation (e.g. cross-channel allocation). However, our research gives counter-intuitive conclusions and indicates that they should put the emphasis on more fine-grained allocation instead of more coarse-grained allocation. The reason may be that the precise-targeting characteristics of programmatic advertising driven by big data can not be fully reflected in coarse-grained levels.

Also, when publishers optimize the multi-level ad inventory allocation, the absolute marginal revenue should be substituted by the comparative marginal revenue as the decision basis. The above analysis has proved that the increasing positive marginal revenue can not support publishers to keep increasing allocation to a channel or platform, since the increasing marginal loss of another channel or platform may not be compensated by it.

7. Conclusion and future work

Ad inventory allocation is the significant decision for publishers to maximize revenues, and also the great concern for the supply-demand balance and advertiser-audience match for PA development. In this paper, we formulate the ad inventory allocation as a three-level optimization problem, which views the channel-level, market-level and platform-level allocation as an inter-dependent joint optimization issue. We establish a hierarchical AIAF model, analyze some desirable theoretical properties, and present a generic solution process. Also, we propose an illustrative instantiation to make further demonstrations of AIAF. Computational experiments are designed to validate the models and analysis, and the results show that AIAF can effectively improve publishers' maximal revenues compared with both single-level allocation strategies and two-level allocation strategies. This paper establishes a preliminary ad inventory allocation framework for publishers in PA markets. It not only provides an open and inclusive context from the angle of the PA ecosystem, but also supports the ad inventory decisions in practice from the angle of individual publishers.

In the ongoing work, we will extend this framework to probe more complicated situations, e.g. uncertainties, competitiveness, budget constraints and so on. Also, we try to explore the problem using ACP (artificial systems + computational experiments + parallel execution) approach (Wang et al., 2008; Wen et al., 2013) to combine both the real and artificial markets, resulting in an efficient allocation and adjustment of the ad inventory allocation in PA markets.

Acknowledgements

We gratefully acknowledge the funding supports from the National Natural Science Foundation of China (#71472174, #61533019, #71702182).

References

- Adikari, S., Dutta, K., 2015. Real time bidding in online digital advertisement. *New Horizons in Design Science: Broadening the Research Agenda*. Springer International Publishing 19-38.
- Agrawal, S., Wang, Z., Ye, Y., 2014. A dynamic near-optimal algorithm for online linear programming. *Oper. Res.* 62 (4), 876–890.
- Balseiro, S., Feldman, J., Mirrokni, V., Muthukrishnan, S., 2014. Yield optimization of display advertising with ad exchange. *Manage. Sci.* 2886–2907.
- Broder, A.Z., 2008. In: *Proceedings of the 2008 ACM conference on Recommender systems*, Lausanne, Switzerland, pp. 1–2.
- Busch, O., 2015. *Programmatic Advertising: The Successful Transformation to*

- Automated, Data-Driven Marketing in Real-Time. Springer Publishing Company Incorporated.
- Cachon, G.P., Fisher, M., 2000. Supply chain inventory management and the value of shared information. *Manage. Sci.* 46 (8), 1032–1048.
- Chahuara, P., Grislain, N., Jauvion, G., 2017. Real-time optimization of web publisher RTB revenues. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 1743–1751.
- Chen, Y.J., 2013. Optimal dynamic auctions for display advertising. *SSRN Electron. J. Clerici, A., Perego, S.*, 2010. Inventory allocation for online graphical display advertising. *Comput. Sci.* 7 (7), 135–143.
- Cramton, P., Shoham, Y., Steinberg, R., 2006. *Combinatorial Auctions*. MIT Press.
- Cui, Y., Zhang, R., Li, W., 2011. Bid landscape forecasting in online ad exchange marketplace. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 71–78.
- Dawson, P., Lamb, M., 2016. Enhanced success with programmatic social advertising. In: Busch, O. (Ed.), *Programmatic Advertising. Management for Professionals*. Springer, Cham.
- Devanur, N.R., Hayes, T.P., 2009. The adwords problem: online keyword matching with budgeted bidders under random permutations. In: *ACM Conference on Electronic Commerce*. ACM, pp. 71–78.
- Feldman, J., Korula, N., Mirrokni, V., 2009. Online ad assignment with free disposal. In: *International Workshop on Internet and Network Economics*. Springer-Verlag, pp. 374–385.
- Feldman, J., Henzinger, M., Korula, N., et al., 2010. Online stochastic packing applied to display ad allocation. *Lecture Notes Comput. Sci* abs/1001.5076:182–194.
- Fernandez-Tapia, J., 2015. *Real-Time Bidding Rules of Thumb: Analytically Optimizing the Programmatic Buying of Ad-inventory*. Social Science Electronic Publishing.
- Fernandez-Tapia, J., Guant, O., Lasry, J.M., 2016. Optimal real-time bidding strategies. *Appl. Math. Res. eXpress* 1–42.
- Ghosh, A., McAfee, P., Papineni, K., 2009. *Bidding for Representative Allocations for Display Advertising*. Internet and Network Economics. Springer, Berlin Heidelberg 208–219.
- Harris, M., Kriebel, C.H., Raviv, A., 1982. Asymmetric information, incentives and intrafirm resource allocation. *Manage. Sci.* 28 (6), 604–620.
- Krieger, A.M., Green, P., 2006. E.A tactical model for resource allocation and its application to advertising budgeting. *Eur. J. Oper. Res.* 170 (3), 935–949.
- Lai, H.C., Shih, W.Y., Huang, J.L., 2017. Predicting traffic of online advertising in real-time bidding systems from perspective of demand-side platforms. In: *IEEE International Conference on Big Data*. IEEE, pp. 3491–3498.
- Li, X., Guan, D., 2014. Programmatic Buying Bidding Strategies with Win Rate and Winning Price Estimation in Real Time Mobile Advertising. *Advances in Knowledge Discovery and Data Mining*. Springer International Publishing 447–460.
- Li, J., Yuan, Y., Zhao, X., Wang, F.-Y., 2017. Research on Information Structure of Programmatic Advertising Markets. In: *The 20th World Congress of the International Federation of Automatic Control (IFAC2017)*, Toulouse, France, 9–14 July 2017.
- Li, J., Ni, X., Yuan, Y., 2018. The reserve price of ad impressions in multi-channel real-time bidding markets. *IEEE Trans. Comput. Social Syst.* 5 (2), 583–592.
- McAfee, R.P., 2011. The design of advertising exchanges. *Rev. Ind. Organ.* 39 (3), 169–185.
- Mostagir, M., 2010. Optimal delivery in display advertising. In: *IEEE Xplore* Communication, Control, and Computing, pp. 577–583.
- Muthukrishnan, S., 2009. Ad exchanges: research issues. In: *International Workshop on Internet and Network Economics*. Springer-Verlag, pp. 1–12.
- Najafi Asadolahi, S., Fridgeirdottir, K., 2014. Cost-per-click pricing for display advertising. *SSRN Electron. J.* 16 (4).
- Perlich, C., Dalessandro, B., Hook, R., Stitelman, O., Raeder, T., Provost, F., 2012. Bid optimizing and inventory scoring in targeted online advertising. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 804–812.
- Qin, R., Yuan, Y., Wang, F.Y., 2017. Exploring the optimal granularity for market segmentation in RTB advertising via computational experiment approach. *Electron. Commer. Res. Appl.* 24.
- Radovanovic, A., Heavlin, W.D., 2012. Risk-aware revenue maximization in display advertising. In: *Proceedings of the 21st International Conference on World Wide Web*. ACM, pp. 91–100.
- Roels, G., Fridgeirdottir, K., 2009. Dynamic revenue management for online display advertising[J]. *J. Revenue Pricing Manage.* 8 (5), 452–466.
- Sabbaghi, N., Sheffi, Y., Tsitsiklis, J.N., 2014. Allocational flexibility in constrained supply chains. *Int. J. Prod. Econ.* 153 (4), 86–94.
- Simpson, R.W., 1989. Using Network Flow Techniques to Find Shadow Prices for Market Demands and Seat. *Flight Transportation Laboratory Memorandum M89-1*.
- Stavrogiannis, L.C., Gerding, E.H., Polukarov, M., 2014. Auction mechanisms for demand-side intermediaries in online advertising exchanges. In: *International Conference on Autonomous Agents and Multiagent Systems*, pp. 1037–1044.
- Sudharshan, D., 1995. *Marketing Strategy: Relationships, Offerings, Timing, and Resource Allocation*. Prentice-Hall.
- Vee, E., Vassilvitskii, S., Shanmugasundaram, J., 2010. Optimal online assignment with forecasts. In: *ACM Conference on Electronic Commerce*. ACM, pp. 109–118.
- Walsh, W.E., Boutilier, C., Sandholm, T., et al., 2010. Automated channel abstraction for advertising auctions. In: *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI*, pp. 887–894.
- Wang, F., Zeng, D., Yuan, Y., 2008. An ACP-based Approach for Complexity Analysis of E-commerce System [J]. *Complex Systems and Complexity Science* 3.
- Wen, D., Yuan, Y., Li, X.R., 2013. Artificial societies, computational experiments, and parallel systems: An investigation on a computational theory for complex socio-economic systems. *IEEE Trans. Serv. Comput.* 6 (2), 177–185.
- Wu, C.H., Yeh, M.Y., Chen, M.S., 2015. Predicting winning price in real time bidding with censored data. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 1305–1314.
- Yang, J., Vee, E., Vassilvitskii, S., et al., 2010. Inventory allocation for online graphical display advertising. *Comput. Sci.* 7 (7), 135–143.
- Yuan, S., Wang, J., Zhao, X., 2013. Real-time bidding for online advertising: measurement and analysis. In: *International Workshop on Data Mining for Online Advertising*. ACM, pp. 1–8.
- Yuan, Y., Wang, F., Li, J., 2014. A survey on real time bidding advertising. In: *IEEE International Conference on Service Operations and Logistics, and Informatics*. IEEE, pp. 418–423.
- Yuan, S., Wang, J., Chen, B., 2014. An empirical study of reserve price optimisation in real-time bidding. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 1897–1906.