

To match or not to match: Economics of cookie matching in online advertising

Arpita Ghosh, Yahoo! Research

Mohammad Mahdian, Google Inc.

R. Preston McAfee, Google Inc.

Sergei Vassilvitskii, Yahoo! Research

Modern online advertising increasingly relies on the availability of user tracking technology called *cookie-matching* to increase efficiency in ad allocations. Web publishers today use this technology to share information about the websites a user has visited, making it possible to target advertisements to users based on their prior history. This begs the question: do publishers (who are competitors for advertising money) always have the incentive to share online information? Intuitive arguments as well as anecdotal evidence suggest that sometimes a premium publisher might suffer from information sharing through an effect called *information leakage*: by sharing user information with the advertiser, the advertiser will be able to target the same user elsewhere on cheaper publishers, leading to a dilution of the value of the supply on the premium publishers.

The goal of this paper is to explore this aspect of online information sharing. We show that when advertisers are homogeneous (i.e., they value the users similarly, up to a constant multiple), *in equilibrium*, the publishers always agree about the benefits of cookie-matching (i.e., either they all benefit, or they all suffer from it). We also analyze a simple model that exhibits how information leakage can help one publisher and harm the other when the advertisers are not homogeneous.

Categories and Subject Descriptors: J.4 [Social and Behavioral Sciences]: Economics

General Terms: Economics, Theory, Algorithms

Additional Key Words and Phrases: Ad Auctions, Cookie-matching, Information in Auctions

1. INTRODUCTION

When will competitors share online information? We consider this question in the context of Internet cookies, which are small files placed on a user's computer that permit a website to record information about a previous visit. Cookies can be used to benefit the user, e.g., automated login and remembering user preferences, and can also be used to target advertising.

Some websites have begun sharing cookie information. This is most noticeable to the user when, for instance, they search for flights to Hawaii on Orbitz.com, and find Hawaii ads following them across many the web, for example showing up on the New York Times website. Sharing of cookies creates some obvious conveniences for the user, and also permits more targeted advertising by creating a more detailed picture of the

Author's email addresses: Arpita Ghosh, arpita@gmail.com; Mohammad Mahdian, mahdian@alum.mit.edu; R. Preston McAfee, preston@mcafee.cc; Sergei Vassilvitskii, sergeiv@gmail.com.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

EC'12, June 4–8, 2012, Valencia, Spain.

Copyright 2012 ACM 978-1-4503-1415-2/12/06...\$10.00.

customer. For example, if Walmart knows a user visited a site focused on infant health, it may advertise baby strollers to the same user. The question we address is when companies will voluntarily agree to participate in cookie-sharing, and what effects on prices and profits cookie-sharing creates.

Sharing cookie information is done through a service, called *cookie matching*, that is currently offered by most online advertising exchange markets. Cookie-matching means that when a user visits a website, the ad exchange scrambles the cookie that they have placed on the user's computer using a collision-resistant one-way hash function, and passes the scrambled cookie to the interested advertisers. This scrambling, while not revealing the contents of the cookie, enables these advertisers to discover if they have interacted with this user before (for example, if the user has visited their website before, or if they have advertised to the user before). The advertisers will then be able to decide how much they want to bid based on this information. This sharing also enables the advertisers (both the winner of the auction and also the losers) to identify the user in his/her future visits (unless the cookie is deleted in the meantime).

The obvious benefit of this mechanism is that it allows advertisers to target users that have previously shown interest in their products. This targeting increases the value of advertising to the advertisers, and some of this increased value will be passed to the publishers in the long run. The not-so-obvious drawback for the publishers, especially premium publishers, is *information leakage*: by passing the scrambled cookie to the advertiser, the advertiser will be able to target the same user elsewhere on cheaper publishers. Information leakage dilutes the value of the supply on the premium publishers.

The goal of this paper is to explore the latter aspect of cookie-matching, i.e., when through cookie-matching, a publisher leaks valuable information that could harm their revenue (while helping other publishers). We start with a discussion of various effects of cookie-matching in ad auctions.

1.1. The impact of cookie-matching on auction revenue

We can divide the effects of providing data through cookie-matching on the ad auction revenue into four major categories:

- **More efficient allocation:** Data (whether it is labeling the impressions with features of the user, or with identifiable information like the scrambled cookie) allows the advertisers to evaluate the impressions more accurately, thereby increasing the efficiency of the allocation.
- **Market fragmentation:** The increased efficiency can lead to the fragmentation of the market and decreased revenue for the auctioneer. An example is when two advertisers are competing for an ad slot, but one advertiser is only interested in male users while the other is interested in females. In this scenario, providing any data that helps advertisers distinguish male and female users will lead to a more efficient allocation, but it can also lead to a lower revenue in a second-price auction.
- **Better interaction with the user:** Cookie matching allows the advertisers to know how many times they have seen a user before, thereby personalizing their ad creative each time and avoiding advertising too many times to the same user. The latter effect is called frequency-capping.
- **Data leakage:** By passing the cookie information to the advertiser, the publisher (and the ad exchange platform) run the risk that the advertiser takes advantage of

this data elsewhere (on different publishers or even on different platforms) to decrease her cost at the cost of decreased revenue to the publisher.

The market fragmentation effect (and its positive counterpart, allocation efficiency) is essentially the reverse of the *bundling* problem that has been studied in the auction theory literature [see McAfee et al. 1989]. In this paper, our focus is on the information leakage effect, and will give a model that demonstrates this effect in the equilibrium, as well as a positive result that shows that for a large class of models, information leakage is not a problem in the equilibrium.

1.2. Behind the scenes of this paper

This work started with the goal of coming up with a simple model that exhibits the information leakage effect. The intuitive explanation, as well as anecdotal evidence suggests that this is a real effect with many practical implications, and therefore we must be able to capture it in the equilibrium of a simple model. To achieve this, we focused on a model that avoids getting into the complexity of the first two effects by assuming a homogeneous set of advertisers, i.e., a set of advertisers that value the users similarly, up to a constant multiple. This is a realistic model in many situations where even without cookie-matching there is already enough background information (e.g., user demographics or impression type) to divide the market into fine-enough segments, each with a homogeneous set of competing advertisers.

To our surprise, after analyzing a simple special case of such a model (presented in Section 2), we found out that in equilibrium, there is always agreement between the publishers, i.e., either they all get a higher revenue, or they all get a lower revenue (depending on whether frequency capping is revenue positive or revenue negative). We then generalized this result to quite a general model (defined in Section 2). Essentially, the only major assumption of our model is the homogeneity of advertisers. The results, presented in Section 3, suggest that the information leakage effect might not be as serious a problem in practice as it might appear. We also give a simple model with only two types of advertisers where this result breaks, i.e., the information leakage causes a disagreement between the publishers.

1.3. Related work

Information sharing by competitors has long been studied by economists. Much of the literature is devoted to auctions, at least since Milgrom and Weber [1982] and most recently in Abraham et al. [2011]. However, this literature has focused on information provided to competitors, rather than shared by competitors. There are a few studies focused on incentives to share by competitors, such as Clarke [1983] and Gal-Or [1985], which conclude that firms will not voluntarily share information. Finally, there is an extensive literature on information-sharing in a cartel environment; see e.g. Teece [1994] and the references therein. Information sharing is often viewed as a sign of collusion on the principle that firms have no incentive to provide information except to produce a cartel. The focus of the cartel papers is not on the incentives to join an information-sharing system but instead on the use of such a system to fix prices.

There is also a number of recent papers on the roll of information and targeting in advertising. Bergemann and Bonatti [2011] and Fu et al. [2012] study the effect of introducing targeting information on the revenue of ad auctions. Emek et al. [2012] and Sheffet and Miltersen [2012] discuss the algorithmic question of designing revenue-optimal signaling schemes in second-price auctions. Babaioff et al. [2012] study optimal mechanisms for selling information.

2. MODEL

Assume there are a number of websites $\{w_1, w_2, \dots\} = W$, and a number of users visiting these websites. A number of advertisers are interested in advertising to these users. An advertiser is interested in users who are likely to purchase a product. We model this by assuming that a user is either a *high* type or a *low* type. Each advertiser has a positive value for a high-type user and zero value for a low-type user. The advertisers cannot observe the type of a user, and can only infer it from the user's behavior. Note that in the most general model, a user's type does not have to be limited to high and low, and there could be multiple types corresponding to different revenue levels the advertiser expects to make. For the majority of this paper, we simplify and focus on only two types, a setting rich enough to capture the essence of the arguments. This assumption is for simplicity only, and we will briefly discuss how the result can be generalized to more than two types.

For the positive results, we will make use of a homogeneity assumption, which states that the value of a user to each advertiser can be written as a product of an advertiser specific value, and type specific value (for example, zero for low-type users and one for high-type users). As we will show in Section 4, breaking this assumption can lead to the information leakage phenomenon.

To model the user visits, we assume that at each time period the user of type $t \in \{H, L\}$ leaves the system with probability q , and with the remaining probability, $1 - q$, chooses to visit website $w \in W$ with probability $p_{t,w}$. We assume that this selection is done independently at every time step. Also, assume that a proportion p_t of users are of type t .

There are N advertisers, with the i 'th one having a value of v_i for advertising to a high-type user for the first time. We assume $v_1 \geq v_2 \geq \dots \geq v_N$. A special case of interest is the case of a fully competitive market, i.e., when v_i 's are all the same value v . We assume the advertiser does not get any additional utility by advertising to the same user more than once.¹

We study this model under two types of information sharing regimes, one with cookie-matching and one without, and compare the revenue of the websites under these regimes. In the model without cookie-matching, on every visit the user appears as a new user to every advertiser (there is no information preserved between visits). In the model with cookie-matching, every time a user visits a website, a scrambled version of the user's cookie is sent to all advertisers, and therefore when an advertiser bids for a user u , she knows the entire sequence of websites u has been to prior to this. We analyze and compare the revenue in both models by computing market equilibrium prices and allocation, i.e., a set of prices and an allocation at which

- every impression for which at least one advertiser has non-zero value is sold, and
- each advertiser weakly prefers the impressions she receives to any other set of impressions.

Sometimes there can be more than one set of prices satisfying the above conditions. For example, if there is only one impression and two bidders with values v_1 and v_2 interested in this impression, then allocating the impression to the higher bidder at any price between v_1 and v_2 satisfies both of the above conditions. In such cases, we study the lowest-price market equilibrium (in this example, the equilibrium at price

¹In reality, the value of advertising to a user more than once does not jump to zero after the first time, but goes to zero more slowly. We focus on this special case to simplify the calculations. It is not hard to generalize the results to the more general model.

v_2), which is a natural generalization of the second price auction prices. Note that it is not a priori obvious that a “lowest-price” equilibrium should exist. Our proof also establishes the existence of such an equilibrium. Alternatively, one can apply the result of Demange, Gale, and Sotomayor [1986] that shows that such a canonical equilibrium exists. The revenue per impression of a website w_i (also called the revenue of publisher i) is the expected price of an impression on this website in such an equilibrium.

A special case. The following special case of the model will be used as a showcase throughout the paper: There are two websites w_1, w_2 , where w_1 is the premium website and w_2 is the non-premium one. Half the users are high-type and the other half are low-type (i.e., $p_H = p_L = 1/2$). A high-type user visits each website with probability $1/2$, whereas a low-quality user only visits w_2 . This means that visiting w_1 is a clear signal that the user is of high-type. Also, in the special case we assume all v_i 's are the same value v .

3. PUBLISHER REVENUE WITH HOMOGENEOUS ADVERTISERS

In this section, we analyze the expected revenue of each publisher in both with and without cookie-matching models when the advertisers are homogeneous. The phenomenon of interest to us is disagreement between different publishers about whether or not to share information. More precisely, we would like to know if there are scenarios where providing cookie-matching increases the revenue of one publisher at the expense of another publisher. Our main result is the following theorem, which proves that in the model with homogeneous advertisers, this will never happen.

THEOREM 3.1. *In both with and without cookie-matching models (with homogeneous advertisers), the expected revenue per impression of a website w is proportional to*

$$\beta_w := \frac{p_H p_{H,w}}{\sum_{t \in \{H, L\}} p_t p_{t,w}}.$$

Therefore, either for all websites w_i , the revenue per impression of w_i in the model with cookie-matching is greater than its revenue per impression in the model without cookie-matching, or the reverse inequality holds for all w_i .

Note that the quantity β_w is the fraction of impressions on w that are from high-type users. Therefore, the above theorem shows that in both models, all websites get the same expected revenue per high-type visitor.

The proof of Theorem 3.1 is presented in Sections 3.1 (for the model without cookie-matching) and 3.2 (for the model with cookie-matching). We discuss further generalizations of this result in Section 3.3. Also, in section 3.4 we numerically examine the calculated revenues in the case of the simple model.

3.1. Analysis of the model without cookie-matching

In the model without tracking, all impressions on a website look the same. Therefore, for each website w , there is a single price θ_w . We now write equilibrium conditions for these prices.

Consider the utility maximization problem from the perspective of one fixed advertiser a . This advertiser needs to decide what fraction $x_{a,w}$ of traffic on each website w to buy. We compute the utility per user that this advertiser derives from this allocation. Fix a user of type $t \in \{H, L\}$. The expected total number of visits of this user is $1/q$.

On each such visit, the probability that the user visits w is $p_{t,w}$, in which case with probability $x_{a,w}$ she sees the ad of a . Therefore, in total, a pays a cost of

$$\frac{1}{q} \sum_{w \in W} p_{t,w} x_{a,w} \theta_w$$

for this user. Also, the probability that the user visits the website exactly i times is $q(1-q)^{i-1}$. So, since in each visit, the probability of seeing a 's ad is $x_{a,t} := \sum_{w \in W} p_{t,w} x_{a,w}$, the probability of getting exposed to this ad at least once can be written as:

$$1 - \sum_{i \geq 1} q(1-q)^{i-1} (1-x_{a,t})^i = \frac{x_{a,t}}{q + (1-q)x_{a,t}}.$$

Therefore, the total utility that a derives from a random user is

$$U = \frac{v_a p_H x_{a,H}}{q + (1-q)x_{a,H}} - \frac{1}{q} \sum_{t \in \{H,L\}} p_t \sum_{w \in W} p_{t,w} x_{a,w} \theta_w.$$

To optimize, we need to take the derivative of the above expression with respect to each $x_{a,w}$:

$$\frac{\partial U}{\partial x_{a,w}} = \frac{v_a p_H p_{H,w} q}{(q + (1-q)x_{a,H})^2} - \sum_{t \in \{H,L\}} \frac{p_t p_{t,w} \theta_w}{q}$$

The above derivative is zero if and only if

$$\theta_w = v_a p_H p_{H,w} \left(1 + \left(\frac{1}{q} - 1\right) x_{a,H}\right)^{-2} \left(\sum_{t \in \{H,L\}} p_t p_{t,w}\right)^{-1}$$

or

$$\theta_w = v_a \left(1 + \left(\frac{1}{q} - 1\right) x_{a,H}\right)^{-2} \beta_w. \quad (1)$$

This means that for every advertiser a and website w , either

- Equation (1) holds; or
- $x_{a,w} = 0$ and $\theta_w > v_a \left(1 + \left(\frac{1}{q} - 1\right) x_{a,H}\right)^{-2} \beta_w$; or
- $x_{a,w} = 1$ and $\theta_w < v_a \left(1 + \left(\frac{1}{q} - 1\right) x_{a,H}\right)^{-2} \beta_w$.

We are now ready to complete the proof of Theorem 3.1 in the case of no cookie-matching. Assume, for contradiction, that at a lowest-price equilibrium, for two websites w and w' , we have $\theta_w/\beta_w > \theta_{w'}/\beta_{w'}$. For any advertiser a , if $x_{a,w} > 0$ and $x_{a,w'} < 1$, we have

$$v_a \left(1 + \left(\frac{1}{q} - 1\right) x_{a,H}\right)^{-2} \geq \theta_w/\beta_w > \theta_{w'}/\beta_{w'} \geq v_a \left(1 + \left(\frac{1}{q} - 1\right) x_{a,H}\right)^{-2}, \quad (2)$$

which is a contradiction. Therefore, for every a , either $x_{a,w} = 0$ or $x_{a,w'} = 1$. Since at the equilibrium there must be at least one a^* with $x_{a^*,w} > 0$. Therefore, for this a^* , we have $x_{a^*,w'} = 1$. This means that no other $a \neq a^*$ can have $x_{a,w'} = 1$, and therefore for all such a 's, we have $x_{a,w} = 0$, implying that $x_{a^*,w} = 1$. Since this argument holds for every

two w, w' with $\theta_w/\beta_w > \theta_{w'}/\beta_{w'}$, we conclude that a^* must have bought everything, i.e., $x_{a^*,w} = 1$ for all w . We can now complete the proof using the fact that the equilibrium is a *lowest-price* equilibrium. Take a website w with the maximum value of θ_w/β_w . It is easy to see that by slightly decreasing the price of impressions at this website, no advertiser $a \neq a^*$ will be interested in buying these impressions, and therefore we are still at an equilibrium, contradicting the minimality assumption.

The contradiction shows that at a lowest-price equilibrium, for every two websites w and w' , we must have $\theta_w/\beta_w = \theta_{w'}/\beta_{w'}$. \square

A closed form in the simple model. We can give a simple closed-form expression for the revenue in the case that all advertisers have the same value v . In this case, due to symmetry, in the equilibrium we must have $x_{a,H} = 1/N$ for every advertiser a . Therefore, Equation (1) implies

$$\theta_w = v \left(1 + \left(\frac{1}{q} - 1\right) \frac{1}{N}\right)^{-2} \beta_w. \quad (3)$$

Plugging in the parameters of the simple model, we get:

$$\theta_1 = \left(1 + \frac{1-q}{qN}\right)^{-2} v \quad \text{and} \quad \theta_2 = \left(1 + \frac{1-q}{qN}\right)^{-2} v/3.$$

3.2. Analysis of the model with cookie-matching

We now analyze the model in the cookie-matching regime. In this case, each impression comes with a complete history of the user (the websites she has visited), and therefore the equilibrium price of advertising to the user can depend on this history. We denote the sequence of websites the user has visited by $\text{hist} = w_{i_1}, w_{i_2}, \dots, w_{i_k}$. The length of this history is denoted by $|\text{hist}| = k$. We denote the history hist appended by a visit to a website $w \in W$ by $\text{hist}.w$. The price at the history hist is denoted by $\lambda(\text{hist})$.

We write the equilibrium conditions for the price at a history hist of length k . At this history, $k-1$ advertisers (that can be shown by induction to be the advertisers $1, \dots, k-1$, i.e., the advertisers with the top $k-1$ values) have already won an impression, and therefore only advertisers k, \dots, N are interested. The value of the i th advertiser for buying this impression is $\Pr[H|\text{hist}] \cdot v_i - \lambda(\text{hist})$, where $\Pr[H|\text{hist}]$ denotes the probability that the user is of type H , given the history of the sites she has visited.

The utility of this advertiser for waiting is $(1-q)$ (the probability that the user returns) times the utility conditioned on her return. If the user returns and visits a website w , the utility of the advertiser is $v_i - \lambda(\text{hist}.w)$ if the user is of high type and $-\lambda(\text{hist}.w)$ if she is of low type. Therefore, the overall utility of the advertiser if the user returns can be written as:

$$\begin{aligned} & \Pr[H|\text{hist}] \cdot \sum_{w \in W} p_{H,w} \cdot (v_i - \lambda(\text{hist}.w)) + \Pr[L|\text{hist}] \cdot \sum_{w \in W} p_{L,w} \cdot (-\lambda(\text{hist}.w)) \\ & = \Pr[H|\text{hist}] v_i - \sum_{w \in W} \Pr[w|\text{hist}] \lambda(\text{hist}.w), \end{aligned} \quad (4)$$

where $\Pr[w|\text{hist}] = \Pr[H|\text{hist}] \cdot p_{H,w} + \Pr[L|\text{hist}] \cdot p_{L,w}$ is the probability that a user visits the website w after the history hist .

Therefore, the equilibrium condition says that for the advertiser who wins the present impression, the buy-now utility of $\Pr[H|\text{hist}].v_i - \lambda(\text{hist})$ is greater than or equal to $(1-q)$ times the expression in (4), and for the other advertisers the reverse inequality holds. This implies that the advertiser winning this impression should be the advertiser k , and in order to get the lowest-price equilibrium, we must have equality for the advertiser $k+1$. Thus, the equilibrium condition implies:

$$\begin{aligned} \Pr[H|\text{hist}].v_{k+1} - \lambda(\text{hist}) = \\ (1-q) \left(\Pr[H|\text{hist}].v_{k+1} - \sum_{w \in W} \Pr[w|\text{hist}]\lambda(\text{hist}.w) \right). \end{aligned} \quad (5)$$

This implies the following recurrence that gives the price at any history in terms of prices at longer histories.

$$\lambda(\text{hist}) = q \cdot \Pr[H|\text{hist}].v_{k+1} + (1-q) \sum_{w \in W} \Pr[w|\text{hist}].\lambda(\text{hist}.w). \quad (6)$$

For the base of this recurrence, we have

$$\forall \text{ hist}, |\text{hist}| \geq N : \lambda(\text{hist}) = 0. \quad (7)$$

This is because after any history that contains at least N page visits, all but at most one of the advertisers have already advertised to the user and therefore the price drops to zero.

Using this recurrence, and induction on the length of the history, we obtain the following formula for the price at any given history:

$$\lambda(\text{hist}) = \Pr[H|\text{hist}]. \sum_{i=k+1}^N q(1-q)^{i-k-1} v_i \quad (8)$$

For each website $w \in W$, the expected revenue per user of this website can be written in terms of $\lambda(\cdot)$'s as follows:

$$\text{Revenue per user of } w = \sum_{\text{hist}=(w_{i_1}, \dots, w_{i_{k-1}}, w)} \Pr[\text{hist}]\lambda(\text{hist}) \quad (9)$$

Using (8) and (9), we can write the expected revenue per user of a website w as follows:

$$\text{Revenue per user of } w = \sum_{k=1}^N \sum_{\text{hist}} \Pr[\text{hist}]. \Pr[H|\text{hist}]. \sum_{i=k+1}^N q(1-q)^{i-k-1} v_i,$$

where the second summation is over all histories hist of length k that end in w . Using the Bayes rule, we have

$$\text{Revenue per user of } w = \sum_{k=1}^N \sum_{\text{hist}} p_H \cdot \Pr[\text{hist}|H] \cdot \sum_{i=k+1}^N q(1-q)^{i-k-1} v_i.$$

The summation $\sum_{\text{hist}} \Pr[\text{hist}|H]$ is equal to the probability that a high-type user visits at least k websites, and on her k 'th visit, chooses the website w . This can be written as $(1-q)^{k-1} p_{H,w}$. Using this, we can simplify the above expression:

$$\begin{aligned} \text{Revenue per user of } w &= \sum_{k=1}^N p_H \cdot (1-q)^{k-1} p_{H,w} \cdot \sum_{i=k+1}^N q(1-q)^{i-k-1} v_i \\ &= p_H p_{H,w} q \sum_{i=2}^N \sum_{k=1}^{i-1} (1-q)^{i-2} v_i \\ &= p_H p_{H,w} q \sum_{i=2}^N (i-1)(1-q)^{i-2} v_i \end{aligned} \quad (10)$$

The summation in the above expression is independent of w . This means that the revenue per user of w is proportional to $p_H p_{H,w}$. Also, note that a random user on each visit chooses w with probability $\sum_{t \in \{H,L\}} p_t p_{t,w}$. Therefore, the expected number of impressions that a user generates on w is proportional to $\sum_{t \in \{H,L\}} p_t p_{t,w}$. Thus, the expected revenue per impression on w is proportional to $p_H p_{H,w} / (\sum_{t \in \{H,L\}} p_t p_{t,w}) = \beta_w$. This completes the proof of Theorem 3.1 in the model with cookie-matching. \square

A closed-form expression for the simple model. In the case that all advertisers have the same value v , the pricing solution (8) can be simplified to

$$\lambda(\text{hist}) = v \cdot \Pr[H|\text{hist}] \cdot (1 - (1-q)^{N-k}). \quad (11)$$

Also, Equation (10) can be simplified as follows:

$$\begin{aligned} \text{Revenue per user of } w &= p_H p_{H,w} q v \sum_{i=0}^{N-1} i(1-q)^{i-1} \\ &= \frac{p_H p_{H,w} v}{q} (1 - N(1-q)^{N-1} + (N-1)(1-q)^N) \end{aligned} \quad (12)$$

In the simple model, a random user in expectation creates $1/q$ impressions, and each impression will be on w_1 with probability $1/4$ and on w_2 with probability $3/4$. Therefore, the expected revenue per impression in this model is $(1 - N(1-q)^{N-1} + (N-1)(1-q)^N)v$ for the website w_1 , and $(1 - N(1-q)^{N-1} + (N-1)(1-q)^N)v/3$ for w_2 .

3.3. Further generalization

We presented the results of this section in a model where there are only two types of users. This assumption can be relaxed, as follows: The user can be of any of the types in a set $\mathcal{T} = \{t_1, t_2, \dots\}$. The fraction of the users of type t is p_t , and users of this type visit website w with probability $p_{t,w}$ in each stage. The value of an advertiser a for a user of type t can be written as $\gamma_t v_a$. Think of γ_t as the conversion rate (i.e., probability

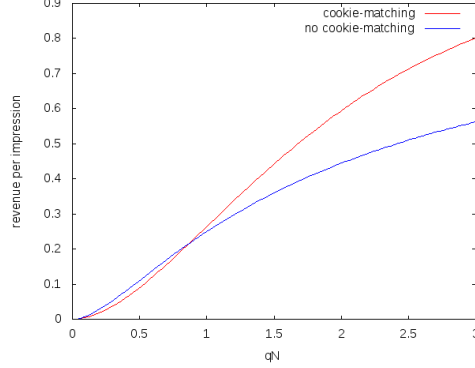


Fig. 1. Revenue comparison between the two scenarios in the simple model

of purchasing the product) of users of type t , and v_a as the profit per conversion for advertiser a . The case of two types H, L corresponds to setting $\gamma_H = 1, \gamma_L = 0$.

We can generalize Theorem 3.1 to this more general model as follows:

THEOREM 3.2. *In both with and without cookie-matching models (with homogeneous advertisers and multiple types of users, as defined above), the expected revenue per impression of a website w is proportional to*

$$\beta_w := \frac{\sum_t p_t p_{t,w} \gamma_t}{\sum_t p_t p_{t,w}}.$$

Therefore, either for all websites w_i , the revenue per impression of w_i in the model with cookie-matching is greater than its revenue per impression in the model without cookie-matching, or the reverse inequality holds for all w_i .

3.4. Numerical examination of the simple model

It is instructive to look at the closed-form expressions for the revenue in the simple model. Setting $v = 1$ and N to be a large number, we have plotted the revenue per impression of w_1 as a function of qN in Figure 1. As can be seen in this figure, there is a range of parameters for which the model without cookie-matching achieves a higher revenue for publishers than the model with cookie-matching. For example, this happens when q is roughly $1/(2N)$, which means that each user visits about $2N$ websites before quitting. The intuitive reason is that in this range, the supply in the cookie matching model ($2N$ impressions per user) is more than the demand (N advertisers), leading to a low price. In the model without cookie-matching, the inefficiency due to the possibility of one advertiser advertising multiple times to the same user artificially decreases the supply, thereby increasing the prices. As the supply ($1/q$ impressions per user) gets more in line with demand (N), the cookie-matching model yields higher revenue for all publishers than the model without cookie-matching.

4. A SCENARIO WITH INFORMATION LEAKAGE

In this Section we show that under heterogeneous advertisers, the impact of cookie matching on different publishers is different, and can lead to a loss of revenue for some and an increase in revenue for others.

Consider the simple model with N advertisers who value high type users at v and low type users at 0 , we will call these type- A advertisers. In addition suppose there

are N_B type B advertisers that have a value of R the first time their ad is shown, *regardless* of the user type.

We will show that in this setting, there exists a setting of N, N_B, R, v , and q , so that w_1 has higher expected revenue (at equilibrium) in the setting without cookie matching, whereas w_2 has higher expected revenue in the setting with cookie matching. This is precisely the information leakage scenario where the premium website w_1 suffers a loss in revenue due to a dilution of its supply of high valued users.

4.1. No cookie matching

From the point of view of type A advertisers, the equilibrium conditions are the same as the ones in Section 3.1. Now consider advertisers of type B . If $N_B \gg \frac{1}{q}$ then advertisers of type B will be willing to pay $R - \epsilon$ per impression, as the chances of a single advertiser seeing the same user twice are nearly 0.

If the prices θ_1, θ_2 and the value of R is such that $\theta_1 > R > \theta_2$ then advertisers of type A will never be allocated any users from w_2 . In this case, $x_{a,H} = \frac{1}{2N}$, as the N type A advertisers evenly split all of the high valued impressions coming to w_1 . By Equation (1), the prices that support such an allocation are: $\theta_1 = v \left(1 + \frac{1-q}{2qN}\right)^{-2}$. Also, at any price greater than $\theta_2 = \theta_1/3$, type A advertisers do not want any of the impressions on w_2 .

We can now compute the revenue on each website.

LEMMA 4.1. *Let $\theta_1 = v \left(1 + \frac{1-q}{2qN}\right)^{-2} = 3\theta_2$. If $\theta_1 > R > \theta_2$, then the per impression revenue to w_1 is exactly θ_1 and the per impression revenue to w_2 is $R - \epsilon$ for an ϵ that tends to zero as N_B tends to infinity.*

4.2. Cookie matching

When cookie matching is enabled, then the price advertisers are willing to pay is determined by the previous history of the user. In particular, any user who has ever visited w_1 is guaranteed to be a type H user, regardless of which website he is currently visiting. Furthermore, if

$$\frac{v}{2^{N+1}} > R, \quad (13)$$

then the expected value of a user who has visited w_2 k times without visiting w_1 to a type A advertiser is $v \cdot Pr[H|hist] = \frac{v}{2^{k+1}} \geq R$. Therefore the expected value to a type A advertiser is larger than the value to a type B advertiser. This implies that the per impression price for visits $N + 1$ and onwards for *any* user will be R .

Let λ_i be the price of the i -th impression of the user with a history containing at least one visit to w_1 . Adapting Equation (5) to this specific setting, with the base case of the recurrence as $\lambda_j = R$ for all $j \geq N$, we get:

$$\lambda_k = v - (1 - q)^{N-k}(v - R) \quad (14)$$

Therefore, the expected revenue per user to w_1 is:

$$\sum_{k=1}^N \frac{(1-q)^{k-1}}{4} (v - (1-q)^{N-k}(v - R)) + \sum_{k=N+1}^{\infty} \frac{(1-q)^{k-1}}{4} R.$$

Since each user creates $1/(4q)$ impressions on w_1 in expectation, the per impression revenue of w_1 is:

$$\begin{aligned}\theta'_1 &= 4q \cdot \sum_{k=1}^N \frac{(1-q)^{k-1}}{4} (v - (1-q)^{N-k}(v-R)) + 4q \cdot \sum_{k=N+1}^{\infty} \frac{(1-q)^{k-1}}{4} R \\ &= v(1 - (1-q)^N) - Nq(v-R)(1-q)^{N-1} + R(1-q)^N\end{aligned}$$

Moreover, the per impression revenue of w_2 strictly increases, since in addition to type B advertisers, type A advertisers also sometimes bid on impressions on w_2 , and whenever they do so, their bid is strictly greater than R .

To demonstrate information leakage, we need to find a setting of R, v, N, q such that satisfy:

$$\begin{aligned}\theta_1 &= v \left(1 + \frac{1-q}{2qN}\right)^{-2} > R > v \left(1 + \frac{1-q}{2qN}\right)^{-2} / 3 = \theta_2 \\ &\frac{v}{2^{N+1}} > R \\ \theta_1 &= v \left(1 + \frac{1-q}{2qN}\right)^{-2} > v(1 - (1-q)^N) - Nq(v-R)(1-q)^{N-1} + R(1-q)^N = \theta'_1\end{aligned}$$

Setting $R = 0.03, v = 1, N = 4$ and $q = 0.05$ entails $\theta_1 = 0.0878, \theta_2 = 0.0293$ and $\theta'_1 = 0.0436$, satisfying the three conditions above. This leads to a lower per impression revenue to the owner of w_1 , and a higher per impression revenue to the owner of w_2 in the cookie matching case.

5. CONCLUSION

Cookie-matching is now commonplace on the internet, with a number of sites sharing cookie information with each other. We investigated the incentives to share cookies and found, to our surprise, that when advertisers have identical rankings of users, publishers agree whether or not to share cookies. That is, either they all want to, or none want to, share.

In both scenarios, advertisers are paying the expected value of the users. With cookie-matching, the expected value is contingent on the user's history. Without cookie-matching, the expectation is taken only over the website. Either way, cookie-matching does not change the nature of visitors to any website, just the knowledge about them. So either that increase in knowledge increases values, in which case publishers unanimously agree that cookie-matching enhances revenues, or lowers values, in which case publishers prefer not to match. Advertiser values might rise or fall because improves the interaction with the user (e.g. permitting frequency capping), which increases values, but identifies a greater supply, decreasing values.

When advertisers disagree about values, cookie-matching may cause data leakage. The simplest example involves two sites and two types of users. Site one attracts only type H users, while site two attracts type H and type L . Some advertisers value only type H and the others are indifferent. With no cookie matching, advertisers buying on site 2 must advertise to both types of users, reducing the willingness to pay. With cookie matching, site 2 can sell some of the type A s separately, increasing the supply of known type A s, reducing the demand to site one. Thus with cookie matching, site one loses revenue, and site two gains.

There is much more to explore in cookie-matching. In particular, we have set aside the bundling aspects of the absence of cookie-matching; cookie-matching leads to market fragmentation. We have not studied the improvement in efficiency associated with cookie-matching. Finally, side-payments might overcome the data leakage problem – would the winners be willing to compensate the losers to buy cookie-matching?

REFERENCES

- ABRAHAM, I., ATHEY, S., BABAIOFF, M., AND GRUBB, M. 2011. Peaches, lemons, and cookies: Designing auction markets with dispersed information. Tech. Rep. MSR-TR-2011-68, Microsoft Research. January.
- BABAIOFF, M., KLEINBERG, R., AND PAES LEME, R. 2012. Optimal mechanisms for selling information. In *Proceedings of the 13th ACM Conference on Electronics Commerce (EC'12)*.
- BERGEMANN, D. AND BONATTI, A. 2011. Targeting in advertising markets: Implications for offline vs. online media. *RAND Journal of Economics* 42, 414–443.
- CLARKE, R. N. 1983. Collusion and the incentives for information sharing. *The Bell Journal of Economics* 14, 2.
- DEMANGE, G., GALE, D., AND SOTOMAYOR, M. 1986. Multi-item auctions. *Journal of Political Economy* 94, 4, 863–72.
- EMEK, Y., FELDMAN, M., GAMZU, I., PAES LEME, R., AND TENNENHOLTZ, M. 2012. Signaling schemes for revenue maximization. In *Proceedings of the 13th ACM Conference on Electronics Commerce (EC'12)*.
- FU, H., JORDAN, P., MAHDIAN, M., NADAV, U., TALGAM-COHEN, I., AND VASSILVITSKII, S. 2012. Ad auctions with data. unpublished manuscript. preliminary version presented in the 7th Workshop on the Economics of Networks, Systems, and Computation (NetEcon'12).
- GAL-OR, E. 1985. Information sharing in oligopoly. *Econometrica* 53, 2.
- MCAFEE, R. P., MCMILLAN, J., AND WHINSTON, M. 1989. Commodity bundling by a monopolist. *Quarterly Journal of Economics*, 371–383.
- MILGROM, P. AND WEBER, R. J. 1982. A theory of auctions and competitive bidding. *Econometrica* 50, 5, 1089–1122.
- SHEFFET, O. AND MILTERSEN, P. B. 2012. Send mixed signals - earn more, work less. In *Proceedings of the 13th ACM Conference on Electronics Commerce (EC'12)*.
- TEECE, D. J. 1994. Information sharing, innovation, and antitrust. *Antitrust Law Journal* 62, 2.