On-Line Profit-Maximization Algorithms for Managing Sponsored Content in Cellular Networks

Dimitris Paparas, Evangelia Skiani
Columbia University, NY, USA,

Yigal Bejerano, Matthew Andrews
Bell-Labs, Alcatel-Lucent, NJ, USA.

***Abstract*—Wireless Service Providers (SPs) typically generate revenue by offering various data plans to their users. Recently, SPs have been exploring another potential source of revenue by allowing *Content Providers* (CPs) to sponsor users' requests for data. With this option, sometimes referred to as *sponsored content*, a CP can provide the requested content to the end user and compensate the SP for the service cost. The user benefits from free content without any charge to her data plan. This work considers an *on-line* setting in which sponsoring decisions are only made when end user requests for content are realized. Our objective is to maximize the combined profit of a given SP both from the users' data plans as well as sponsored content. We consider the commonly offered data plans of *fixed-quota*. In the fixed-quota data plan, a user pays a fixed fee for the right to obtain a pre-determined amount of unsponsored content during each billing cycle (month). We show that for this data plan there is no on-line algorithm with bounded competitive ratio. However, by relaxing the requirements, an efficient on-line algorithm is introduced. Our simulations show the benefits of the proposed algorithm, which outperforms alternative solutions and provides near optimal profit for the SPs for various settings.**

I. Introduction

Recent years have seen an explosion in the use of *mobile data* connections. This has been driven by an increased adoption of smartphones and tablets whose users expect to be able to use mobile applications at any time. Mobile data enables wireless service providers (SPs) to extend their revenue by offering *data plans* to their users. One recent trend is that operators are terminating unlimited data plans and requiring end users to purchase data in fixed monthly quotas, via a *fixed-quota data plan*. One feature of such plans is that revenue is generated solely from the end users. It is well-known in the industry that most end users have certain limits to payment beyond which it is difficult to extract further revenue. When the end users reach those limits they will prefer to stop data consumption (or switch to another SP) rather than pay more.

However, there are many data transactions for which the *content provider* (CP) also derives economic value from providing the content, e.g., via embedded advertisements. In such cases, it is in the interest of the content provider to allow an end user to access its content, even if that end user has already exhausted its monthly data quota. However, when this happens the service provider will require compensation for the extra bandwidth usage that occurs. This gives rise to the notion of *sponsored content*, which is sometimes known as two-sided pricing. In a sponsored content setup, a content provider pays a certain fee to the service provider so that the end user's data quota is not decremented after a visit to the content provider's content. Moreover, the end user can still access the content even when its quota is exhausted. Two instantiations of sponsored content were recently announced. In late 2013 the Korean operator SKT advertised a scheme in which users to the shopping site m.gsshop.com will have their bandwidth charges paid by the content provider [8]. Then, at the Consumer Electronics Show in January 2014, AT&T announced a service that allows any content provider to sponsor the content being transmitted to its end users [4].

Addressing the emerging need to manage sponsored content, we explore in this study *algorithmic solutions for maximizing the combined profit of wireless operators from both sponsored content as well as the data plans offered to end users.* Previous work (e.g. [1]) has analyzed sponsored content in an *offline* setting where sponsoring decisions are made before requests for traffic arrive. *The goal of this paper is to propose a model for how sponsored content could work in an online setting where sponsoring decisions are made in real-time as traffic requests arrive.*

A. Motivating Example

We now consider a particular instantiation of the sponsored content paradigm that motivates the online model that we consider in this paper. In our example an SP provides a service to users where they can specify music videos that they wish to watch. These videos can potentially be provided by many different CPs. If a user watches the music video from a particular CP, then the CP benefits because they are able to show the user a pre-roll advertisement.

When a user request for a music video arrives at the SP, the SP asks each CP whether they are willing to sponsor the request (so that it is free to the end user). If at least one CP is willing then the SP decides which CP will provide the music video to the end user (and the user does not have to pay for the wireless bandwidth that is consumed by the video). If no CP is willing to sponsor the request then the SP directs some default CP to provide the music video but the user has to "pay" for the associated bandwidth out of its own user quota.

We note that in the above setup all the parties (i.e. the SP, CPs and end users) are motivated to participate. CPs are motivated to sponsor since by doing so they are likely to receive more traffic and show more advertisements. The SP is motivated to offer a sponsored content service since it now has a secondary revenue stream from the CPs. It does not solely have to rely on revenue from the end users. Obviously, in such a setting the end-users benefit from free content without being charged for the service.

Our study considers a generic content sponsoring model which captures several practical realizations of sponsored mobile content. For instance, in one realization the desired content, e.g., a music video, can be specified by its URL, which explicitly specifies the serving CP. In this case the only decision that the CP/SP have to make is whether or not the bandwidth associated with the video is sponsored. There is no notion of selection among different CPs. Another realization may involve a Mobile App that is provided by the SP and allows the user to select desired content. The app then sends the request to the SP and allows it to check sponsoring options among the different CPs.

B. Our results

We focus on an online setting in which user requests for content arrive in an online manner and *multiple* content providers can bid to sponsor these requests. The service provider then decides for each request whether or not sponsoring should take place. In this study the sponsoring decision must be made *online* without any knowledge of future requests (i.e. we do not assume any predictable behavior of the users). If sponsoring should take place the service provider decides which content provider should do it. Content providers cannot make unlimited bids. Each content provider has a *budget* that it can use for sponsoring content. To the best of our knowledge, we are the first to propose an *online optimization framework* for the challenge of *maximizing the combined profit of wireless service providers from sponsored content as well as users' data plans.* Besides the bids of the content providers, our solutions also take into account the service cost for the operator to deliver the request to the user over the wireless network.

As we discuss in the next section, our problem bears some resemblance to the Sponsored Search problem in which search engines wish to sell sponsored links in response to search queries. Our main goal is to investigate whether the ideas from Sponsored Search can be carried over to the Sponsored Content setting. Our contribution can be summarized as follows:

In Section III, we present a concrete model and problem statement for the case in which sponsoring decisions are made in an online fashion. We consider the commonly used *Fixed-Quota* data plan model in which each end user pays in advance for a fixed data quota for unsponsored content.

In Section IV, we show that the Fixed-Quota profit maximization problem has elements of both the Adwords [11], [3] and the Online Generalized Assignment Problem (GAP) problem [15] depending on whether the end users consume all their quota. We first show that we cannot hope for an algorithm that always has finite competitive ratio with respect to the optimal profit. We then present an approximation algorithm that combines an Adwords algorithm with a GAP algorithm so that we are always competitive with respect to sponsoring revenue for the users that do not consume all their quota and we are always competitive with respect to profit for the users that do consume all their quota.

In Section V, we present a numerical example to show that our Fixed-Quota solution is competitive with the best of the Adwords and GAP algorithms, even as the "correct" approach changes depending on whether an end user's quota is fully depleted in the optimal solution.

II. Related Work

Sponsored Search: In our formulation of the sponsored content problem, we assume that requests arrive over time and sponsoring decisions have to be made in an online fashion. Another related online problem that has received a great deal of attention in recent years is "Sponsored Search" [6], [9], [10]. As already discussed, one of the goals of our paper is to show how techniques from Sponsored Search can be applied to online Sponsored Content. We remark however that our problem is different from Sponsored Search due to two main factors. First, in sponsored search all the revenue comes from sponsoring, there is no notion of revenue from end users. Second, as a consequence of this Sponsored Search has no notion of end user quotas. One of the main challenges in our work is correctly incorporating the concept of end user quotas.

Sponsored Content Analysis: Papers [7], [14] examine a formulation of sponsored content based on network utility maximization in which a service provider sets bandwidth prices to both the end user and the content provider who then jointly determine the data rate so as to maximize an aggregate utility. Other papers [5], [12], [13] study sponsored content in relation to different notions of *net neutrality*. Paper [1] focused on an *off-line* setting where the content provider has to decide in advance how much content it is willing to sponsor during an upcoming month. Accesses to the content are charged against this amount until it is depleted at which point end users have to revert to paying for access.

III. Model and Problem Statement

A. System Model

We consider a sponsored content management system from the viewpoint of a single service provider SP. The system contains n end users (EUs) denoted as $U_1, \dots U_n$ and m content providers (CPs) denoted as $CP_1, \dots, CP_m$. Each content provider has a budget B_j that can be used for sponsoring users' requests, while each user is associated with a payment plan, referred to as a *data plan*, for obtaining unsponsored content. We examine the system dynamics over a fixed time period, e.g. a month,[1] assuming that all the CPs' budgets and the users' data plans are initialized at the beginning of the period.

User requests: Each user U_k makes a series of requests for content $q_{k1}, q_{k2}, \dots$ that arrive over time and we assume that each request can be satisfied by some CPs. The *content size* of a request q_{ki} is denoted by s_{ki} and if a request q_{ki} is served then the SP incurs a *service cost* for sending the requested content to the user. For simplifying our discussion and w.l.o.g. we will assume that the service cost of sending one unit of data is normalized to 1. So s_{ki} denotes both the content size as well as the service cost of the request q_{ki}.

Sponsored content: For each request q_{ki} there is a set[2] ϕ_{ki} of CPs that are willing to sponsor the request and each CP_j, $j \in \phi_{ki}$, places a bid b_{jki}. We assume each bid has to cover at least the service cost, i.e., $b_{jki} \geq s_{ki}$. If the SP chooses[3] a CP_j, $j \in \phi_{ki}$, to sponsor request q_{ki} (and deliver the requested content to user U_k) then CP_j's budget is decremented by b_{jki} and the user's data plan is not charged. The SP incurs a service cost of s_{ki} for carrying the requested content, however, it earns

[1] This ignores the fact that in many countries (e.g. the US) user billing periods are not synchronized. In other countries (e.g. South Korea), user billing periods are indeed synchronized to the calendar month.

[2] The set ϕ_{ki} may be empty, i.e., no CP wishes to sponsor the request.

[3] Note that the SP is allowed to not choose any CP to sponsor.

Symbol	Semantics
m	The number of content providers (CPs).
n	The number of end users.
B_j	The budget of CP_j for sponsoring content.
F_k	The monthly fee of end user U_k.
Q_k	The monthly data quota of user U_k.
q_{ki}	the i-th request of end user U_k.
s_{ki}	The service cost of the request q_{ki}.
ϕ_{ki}	The set of CPs that bid on the request q_{ki}.
b_{jki}	The bid of CP_j, $j \in \phi_{ki}$ on the request q_{ki}.
x_{jki}	$= 1$ if request q_{ki} is sponsored by CP j.
y_{ki}	$= 1$ if request q_{ki} is served, 0 otherwise.

TABLE I: Main notation used in this paper.

a revenue of b_{jki}. Hence, the SP's profit from this transaction is $b_{jki} - s_{ki}$. To keep track of sponsored content, we define a variable x_{jki} for every request q_{ki} and CP_j. We set $x_{jki} = 1$ if q_{ki} is sponsored by CP_j and $x_{jki} = 0$, otherwise.

Data Plans: For supporting its operation cost, we assume that the SP provides a *Fixed-quota* plan to each of its users. In this plan, each end user U_k pays a fixed monthly fee of F_k for the right to access a limited amount of at most Q_k units of unsponsored content. We assume $F_k > Q_k$ to ensure that the SP obtains a strictly positive profit from EU U_k.

A request q_{ki} of user U_k that is not sponsored is served if and only if the user has sufficient unused quota. On the other hand, a request that can be sponsored may be served even if the user does not have enough quota.[4] For every served request q_{ki} the quota of user U_k is reduced by s_{ki}. Moreover, the SP incurs a cost of s_{ki} for delivering the content, so its profit from this transaction is $-s_{ki}$. Since a request q_{ki} of user U_k may not be served at all, in addition to the variables x_{jki}, which are defined above, we let $y_{ki} \in \{0, 1\}$ denote whether or not request q_{ki} is served at all. From the above discussion we have that $y_{ki} \geq \sum_j x_{jki}$.

Notation: In certain sections we focus on a single CP and/or a single end user. For these cases we typically omit the parameters j and/or k as appropriate.

B. Problem Statement

We focus on the SP goal to maximize its total profit, which equals the revenue from the data plans and content sponsoring minus the total service cost of carrying the requested content. In the **Fixed-quota (FQ)** data plan the SP profits both from *sponsored content as well as unused users' quotas.* The total profit from sponsoring is $\sum_{jki}(b_{jki} - s_{ki}) \cdot x_{jki}$, while the net-profit from unsponsored requests is $\sum_k F_k - \sum_{ki} s_{ki} \cdot (y_{kr} - \sum_j x_{jkr})$. Notice that $(y_{kr} - \sum_j x_{jkr})$ is 1 only if request q_{ki} is served but not sponsored, otherwise this expression equals 0. We now provide a mathematical definition of our objective in the form of an integer linear program (ILP). (Since the ILP is defined by future requests it cannot be used to actually compute a solution. More generally, we work in the framework of competitive analysis and so we have no prior information (stochastic or otherwise) regarding future requests.)

[4] In practice, a user that has exhausted its data quota may purchase additional quota. Such an option adds additional difficulty to the optimization and we defer it to future work.

$$
\begin{aligned}
P_{OPT} \quad = \quad & \max \left\{ \sum_k F_k + \sum_{jki} b_{jki} \cdot x_{jki} - \sum_{ki} s_{ki} \cdot y_{ki} \right\} \\
\text{s.t.} \quad & \sum_{j \in \phi_{ki}} x_{jki} \leq 1 \quad \forall q_{ki} \\
& \sum_{ki} b_{jki} \cdot x_{jki} \leq B_j \quad \forall j \\
& \sum_i (y_{ki} - \sum_j x_{jki}) \cdot s_{ki} \leq Q_k \quad \forall k \\
& y_{ki} \geq \frac{Q_k - s_{ki} - \sum_{r=1}^{i-1} s_{kr}(y_{kr} - \sum_j x_{jkr})}{Q_k} \quad \forall k \\
& x_{jki} \in \{0, 1\} \ \forall j \in \phi_{ki}, \quad x_{jki} = 0 \ \forall j \notin \phi_{ki}
\end{aligned}
$$

The above formulation of the objective comes from the fact that, besides the fixed revenue from the monthly user payments $\sum_k F_k$, the total revenue from sponsoring is $\sum_{jki} b_{jki} \cdot x_{jki}$ and the total service cost of delivering the content is $\sum_{ki} s_{ki} \cdot y_{ki}$. The first constraint ensures that each request q_{ki} is sponsored by at most one of the CPs in ϕ_{ki}, while the second constraint captures the notion that CP_j cannot spend more than its budget B_j on sponsoring. The following two constraints put restrictions on the variables y_{ki}. The third constraint ensures that the amount of unsponsored content which is served to user U_k is at most its data quota Q_k, while the fourth condition guarantees that the system serves request q_{ki} if user U_k has sufficient data quota. Note that $x_{jki}, y_{ik} \in \{0, 1\}$.

C. Online Competitive Algorithmic Tools

Unlike offline optimization problems, in the case of an online problem the algorithm must decide how to respond to each incoming request without knowledge of the future input. An online algorithm is termed α-competitive if its performance is at most α times worse than the optimal offline algorithm for any possible input [2]. We now describe two online problems that we make use of in our study. Both of these problems have been applied in the context of Sponsored Search.

The AdWords Problem [11], [3]: Consider a system with m bidders that bid on a set W of query words. Each bidder i has a fixed budget B_i for a given time period and it specifies a bid b_{iw} for every word in W. The system gets as input a sequence of requests (also termed queries), one at a time, and it may assign a request q to bidder i for profit of b_{iw}. The objective is to maximize the system profit while preserving the bidder budget constraints. In [11] Mehta et al. introduce a simple online algorithm with competitive ratio of $1 - (1/e)$, under the assumption that bids are small compared to the budgets. To achieve this competitive ratio the algorithm balances between the bid values and the unspent budget of the bidders. Let $\psi(x) = 1 - e^{x-1}$ and let r_i, $0 \leq r_i \leq 1$, be the relative portion of budget currently spent by bidder i. The algorithm allocates the next query w to bidder i with the maximal product $b_{iw} \times \psi(r_i)$. In [3] the authors achieved the same competitive ratio by using a primal-dual approach.

The Generalized Assignment Problem (GAP) [15]: The GAP problem assumes a system with m bins each with capacity B_j, $j \in \{1, \ldots, m\}$. There is also a set of items

$S = \{I_1, \ldots, I_m\}$, where item I_i has a value v_{ji} and a weight (or alternatively a size) w_{ji} when placed in bin j. A sequence Q of items from S arrive one at a time and must be either placed in one of the bins or discarded. The objective is to maximize the value of the items placed in the bins while respecting the capacity of each bin. When the system has a single bin this problem is known as the *online Knapsack problem*. Assuming that the weight and the value of each item is substantially smaller than the bin sizes, Zhou *et al.* [15] present $1/(\ln(U/L)+1)$-competitive algorithms for the Online Knapsack and GAP problems, where U and L are given upper and lower bounds on the ratio $\frac{v_{ji}}{w_{ji}}$ for every bin j and item i.

To obtain this competitive ratio, the algorithm follows an *optimistic approach*, which expects the arrival of high profit items, i.e., it admits a low profit item into a bin only if a small portion of the bin capacity is used and it reserves most of the bin capacity for high profit items. The algorithm works as follows; Let z_j denote the used portion of bin j. When an item i arrives, it can be admitted to bin j if its *profit ratio* $r_{ji} = \frac{v_{ji}}{w_{ji}} > \Psi_j(z_j)$, for a given function $\Psi_j(z)$. If an item can be admitted to several bins then the bin j with the maximal profit ratio r_{ji} is selected. By selecting for each bin j, $\Psi_j(z) = (U_j \cdot e/L_j)^z \cdot (L_j/e)$, where U_j and L_j are the upper and lower bounds of the profit ratio of any item inserted to bin j, the aforementioned competitive ratio is achieved[5].

IV. The Fixed-Quota Data Plan

In this section, we propose an algorithm for scenarios with a fixed-quota data plan and we start with some observations.

Consider the simple case of a single CP and a single EU (in which case we can drop the dependence on j and k). There exists a phase transition which occurs when the user exceeds the quota. In particular, from this point and onward, the additional profit depends on the difference between the bid size and the associated cost, rather than the bid itself. Furthermore, after the quota is exhausted, the algorithm may opt to refuse to sponsor a request without any penalty.

• If in the optimal solution the user does not spend all her quota then all requests must be served. Hence, the cost of serving a request cannot be avoided and the optimization problem reduces to an *Adwords problem* [11] where the values of the items are the bids $b_i, i = 1, 2, \ldots$ and the budget has size B. In this case of a single CP, the optimal solution is to accept as many bids as possible until the budget is exhausted.

• The problem becomes fundamentally different when the optimal solution reaches the user's quota. In this case, one must sponsor the requests that generate the highest revenue, which is the size of the bid after we subtract the service cost associated with the specific request. Hence, this is now an *online knapsack problem* [15] (i.e, GAP with a single CP) where request i has size b_i and value $b_i - s_i$.

The central idea of our analysis is that there exists a clean dichotomy between the situations mentioned above. On the one hand, if we know in advance that the user will not spend the quota, the best strategy is to sponsor those requests based solely on the bid sizes, since the algorithm will have to pay for the cost of all the requests. In this case the problem reduces to an Adwords problem. On the other hand, if the quota is exceeded, the algorithm's behaviour will change since now there are two options: the option of sponsoring now versus the one of denying service in order to save the budget for later requests that may generate higher revenue. Hence in this case we wish to solve an online knapsack problem.

[5]We slightly modified the algorithm in [15] by considering upper and lower bounds per bin to improve performance without affecting approximation ratios.

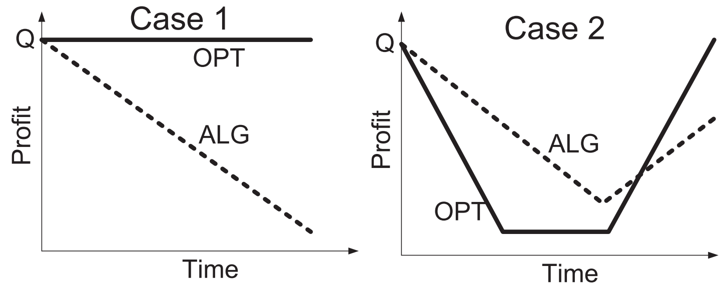

Fig. 1: Example 1 - Case 1 . (Left), Case 2 (Right).

A. Lower Bound on Traditional Competitive Ratio

Ideally we would like an algorithm that has a finite competitive ratio with respect to the profit of the optimal algorithm. Unfortunately, we show in this section that such an algorithm cannot exist. Hence we will consider slightly modified notions of competitive ratio (which we introduce after presenting our bad example.)

Example 1: **Bad Example –** We now present an example to show that for any finite $\varepsilon > 0$, there is no $100\varepsilon/9$-competitive algorithm with respect to SP profit. The example consists of a single CP, a single EU and two types of requests. Type 1 requests have $b_i = 1$ and $s_i = 1$. Type 2 requests have $b_i = 10$ and $s_i = 1$. For a fixed quota size Q, we let $F = Q/(1-\varepsilon)$ and $B = F/\varepsilon$. The adversary begins by sending B requests of type 1. There are two cases to consider (as shown in Figure 1).

• **Case 1:** (Fig. 1 - left.) The online algorithm ALG sponsors at most $(1-\varepsilon)B$ of the Type 1 requests. In this case the adversary does not introduce any more requests and so the optimal solution is to sponsor all requests. (The budget B is large enough to do this.) The optimal profit is therefore $F + \sum_i (b_i - s_i) = F + B(1-1) = F$, whereas the profit of the ALG is at most $F + \sum_i (b_i x_i - s_i y_i)$ which is bounded by $F + (1-\varepsilon)B(1-1) - \min\{Q, \varepsilon B\} \le F - \min\{(1-\varepsilon)F, \varepsilon F/\varepsilon\} = \varepsilon F$.

• **Case 2:** (Fig. 1 - right.) The online algorithm ALG sponsors more than $(1-\varepsilon)B$ of the Type 1 requests. In this case the adversary follows the Type 1 requests with $B/10$ Type 2 requests in which case the optimal solution is to sponsor none of the Type 1 requests. (In the optimal solution note that only Q of the Type 1 requests will actually be served since after that the EU quota is exhausted). Hence the optimal profit is $F + \sum_{i \in \text{Type 2}} (b_i - s_i) - Q$ which equals $F + \frac{B}{10}(10-1) - Q = \varepsilon F + \frac{9B}{10}$. The profit of the ALG is at most $F + \sum_{i \in \text{Type 2}} x_i(b_i - s_i)$ which equals $F + \varepsilon B(10-1) = \varepsilon B + 9\varepsilon B = 10\epsilon B$.

In both cases the competitive ratio of ALG is at most $100\varepsilon/9$.

In light of this lower bound we consider a slightly different notion of competitive ratio. In particular, we divide the EUs into two groups; those for which the unsponsored requests in OPT are below the quota Q_k and those for which the unsponsored requests in OPT are above the quota Q_k. Recall that for the first group we want the sponsoring decisions to be competitive with respect to the actual profit $b_{jki} - s_{ki}$ for a request. For the second group we want to be competitive with respect to the sponsoring revenue b_{jki} since the request will be served regardless. Hence we present algorithms that are competitive with respect to $\sum_{ji}(b_{jki} - s_{ki})$ for the first group of EUs and are competitive with respect to $\sum_{ji} b_{ji}$ for the second group of EUs. The details will be given in the exact statements of Theorems 1 and 2.

We turn to present algorithms for the fixed quota problem that work by combining an algorithm for the Adwords problem with an algorithm for the online GAP problem. In particular we assume access to a χ_{aw}-competitive algorithm for the Adwords problem and a χ_{gap}-competitive algorithm for the online GAP problem. For example, if we use the algorithm of Mehta et al. [11] for Adwords and the algorithm of Zhou et al. [15] for online GAP then we can take $\chi_{aw} = 1 - \frac{1}{e}$ and $\chi_{gap} = 1/(\log(U/L) + 2)$ as long as $L \leq (b_{jki} - s_{ki})/b_{jki} \leq U$ and the b_{jki} values are small relative to B_j.

B. The Algorithm

We now present a randomized algorithm for the fixed quota plan which we denote by FQ$_{ON}$. As requests arrive in an online fashion, the algorithm needs to decide whether to sponsor or not. If the algorithm decides to sponsor, the revenue increases by the bid size minus the service cost; if not, the strategy depends on the status of the user. If the user is "within" quota, the algorithm cannot reject any request, and the revenue must decrease by the service cost. If the user has exceeded the quota, any nonsponsored request is rejected and in that case, the revenue remains the same.

The randomized algorithm, FQ$_{ON}$, virtually runs in parallel the following two algorithms:

1) A_1: Online generalized assignment with m bins. The jth bin has size B_j. If request q_{ki} is placed in bin j this has both value and size b_{jki}. (Note that this is equivalent to the Adwords problem and so we assume a competitive ratio of $\chi_{aw} = 1 - (1/e)$.)

2) A_2: Online generalized assignment with m bins. The jth bin has size B_j. If request q_{ki} is placed in bin j this has value $b_{jki} - s_{ki}$ and size b_{jki}.

Every time a new request q_{ki} arrives FQ$_{ON}$ flips a fair coin and follows one of the above algorithms.

We remark that in order to for our analysis to hold, each algorithm A_1 and A_2 is run *in the background*. That is, we record the entire state of the system as if both A_1 and A_2 were being run on the input. In particular, each algorithm A_i makes its decisions as if the current quota and budget balances were calculated according the decisions of A_i. It does *not* base its decisions on the current quota and budget balances of the combined randomized algorithm.

C. Example

We now present a toy example that illustrate how FQ$_{ON}$ operates. Later, in Section V we describe the performance of the algorithm numerically on a more complex scenario. Our toy example is motivated by the bad example described in Section IV-A and it assumes two CPs, two end users and two request types both with the same service cost of 1. The budgets and quotas of all CPs and EUs equal $B = Q$. The monthly service fee $F = Q + \varepsilon$ for both EUs. The two CPs have the same bid of 1 on Type 1 requests , i.e, $b_{11} = b_{21} = 1$, but different bids on Type 2 requests, $b_{12} = 2$ and $b_{22} = 10$. Both users start with sending B Type 1 requests, which terminate at the same time, follow by a sequence (which may be empty) of Type 2 requests.

Algorithm A_1: Note that Algorithm A_1 will simply sponsor all of the requests until the CPs' balance is exhausted. We consider the following two cases.
• *Case 1:* The EUs send only Type 1 requests. All these requests are sponsored and the profit is simply the two service fees, i.e. $2F$.
• *Case 2:* Each EU sends also a long enough sequence of Type 2 requests, these requests are charged to the user's quota until it is used up and consequently A_1's profit is $2(F - Q) = 2\varepsilon$.

Algorithm A_2: This algorithm sponsors only requests with bid value higher than the service cost. Thus, it does not sponsor any Type 1 requests.
• *Case 1:* The EUs send only Type 1 requests which will be charged to the EU quotas. Thus, the profit is $2(F - Q) = 2\varepsilon$.
• *Case 2:* The CP budgets will be used to sponsor the Type 2 requests (first the budget of CP 2 is used and then the one of CP 1). So the profit is $2(F - Q) + \frac{B}{2} + \frac{9B}{10} \geq \frac{14}{10} \cdot B$.

Algorithm FQ$_{ON}$: We now illustrate how FQ$_{ON}$ maintains a balance between the two types of good behavior.
• *Case 1:* The algorithm sponsors half of the Type 1 requests and charges the users' quotas for the second half so the profit is $2(F - \frac{1}{2} \cdot Q) = F + \varepsilon$, which is more than half of the optimum profit.
• *Case 2:* Half of the Type 1 requests are sponsored while the other half of the CPs' budgets are used for sponsoring the Type 2 requests. The EUs' quotas are used for delivering the unsponsored requests until they are exhausted. Consequently, the profit in this case is $2(F - Q) + \frac{1}{2}(\frac{B}{2} + \frac{9 \cdot B}{10}) \geq \frac{7}{10} \cdot B$, which is also more than half of the optimal profit.

D. Analysis of FQ$_{ON}$

It follows easily that the total budget for CP j spent will not exceed B_j in expectation, since the expected budget for CP j used by both A_1 and A_2 is at most $B_j/2$. Let κ_{ki} be an indicator random variable which takes value 0 if A_1 is used for request q_{ki} and 1 if A_2 is used. The algorithm outputs a set S of sponsored requests that depends on the sequence κ_i.

Let C_{OPT} be the set of EUs whose quota is exceeded by OPT and let C_{ALG} be the set of EUs whose quota is exceeded by FQ$_{ON}$. Let S_{OPT} be the set of sponsoring triples jki decided by OPT. (Note that there can be at most one such j for each ki.) Let S_{ALG} be the set of sponsoring triples decided by FQ$_{ON}$.

Let $P_{OPT}(C)$ be the profit that OPT achieves with respect to the EUs in set C. Note that $P_{OPT} = P_{OPT}(C_{OPT}) +$

$P_{OPT}(\bar{C}_{OPT})$. We begin with the special case of a single CP and a single EU since it is easier to state.

Theorem 1: Assume a single CP and a single EU,

- If OPT exceeds the quota Q then $\mathbb{E}[P_{ALG}] \geq \chi_{gap} P_{OPT}/2$.
- If OPT does not exceed the quota Q then $\mathbb{E}[P_{ALG}] \geq P_{OPT} - \frac{1}{2}\sum_{i \in S_{OPT}} b_i$.

To see why the bound in the second case is desirable recall that if OPT does not exceed the quota then OPT essentially tries to maximize $\sum_{i \in S_{OPT}} b_i$, (subject to $\sum_{i \in \bar{S}_{OPT}} s_i \leq Q$). The expected profit of FQ$_{ON}$ differs from P_{OPT} by at most half this amount.

Theorem 2: Consider the general situation with multiple CPs and EUs. We can write $\mathbb{E}[P_{ALG}] = P_1 + P_2$ where,

$$P_2 \geq \chi_{gap} P_{OPT}(C_{OPT}),$$
$$P_1 \geq P_{OPT}(\bar{C}_{OPT}) - (1 - \chi_{aw}) \sum_{k \in \bar{C}_{OPT}} \sum_{jki \in S_{OPT}} b_{jki}.$$

Hence one of the following two inequalities must hold:

$$\mathbb{E}[P_{ALG}] \geq \chi_{gap} P_{OPT}/2,$$
$$\mathbb{E}[P_{ALG}] \geq (P_{OPT}/2) - (1 - \chi_{aw}) \sum_{k \in \bar{C}_{OPT}} \sum_{jki \in S_{OPT}} b_{jki}.$$

The proofs are deferred to the Appendix due to space constraints. However, the intution for why they hold are given by the description in Section IV-C. In all situations one of the algorithms has performance close to P_{OPT} and FQ$_{ON}$ makes the same decision as that algorithm for half the requests. In the Appendix we also show that we can derandomize FQ$_{ON}$ by giving both A_1 and A_2 a budget of size $B_j/2$ for CP_j (which ensures that the CP budgets are always respected). The derandomized version of FQ$_{ON}$ sponsors any request that is sponsored by at least one of A_1 or A_2.

V. Numerical Evaluation

Due to space limit, we present one numerical example. Here, we assume there are 2 different end users, multiple CPs and two request types. All requests have $s = 0.1$. Each CP has budget 20 and bids $b = 0.11$ for Type 1 requests and $b = 1$ for Type 2 requests. For each EU the quota is $Q = 50$ and inexpensive requests arrive twice as frequently as lucrative ones. The exact arrival rates are such that for EU 1 the total size of all arrivals is $Q/2$ and for EU 2 the total size of all arrivals is $2Q$. Fig. 2 compares the performance of FQ$_{ON}$ with its constituent algorithms Adwords and online GAP (GAP$_{ON}$) (which do not take into account the fact that users can change their status by exhausting their quota). When the number of CPs is large (and hence quotas are more likely to be exhausted) it is better to follow the Adwords algorithm whereas for fewer CPs, GAP$_{ON}$ is more appropriate. Our algorithm FQ$_{ON}$ lies in the middle and hence can handle either case without a drastic dropoff in performance. This is useful when the budget changes dynamically. For example, in a real-word situation, new bidders are usually added or some existing ones are often removed. In such a case, our algorithm guarantees a good performance throughout, while if we were to use one of the other two, we should anticipate a larger fluctuation in the generated profit.

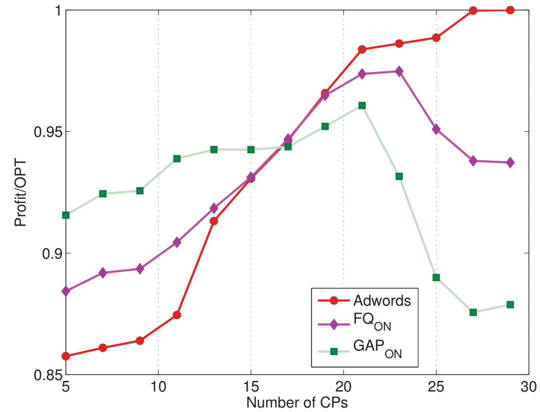

Fig. 2: Ratio of the profit of Adwords, FQ$_{ON}$ and GAP$_{ON}$ algorithms to OPT for Multiple EUs and Multiple CPs.

References

[1] M. Andrews, U. Özen, M. I. Reiman, and Q. Wang. Economic models of sponsored content in wireless networks with uncertain demand. In *INFOCOM*, pages 3213–3218, 2013.

[2] A. Borodin and R. El-Yaniv. *Online Computation and Competitive Analysis*. Cambridge University Press, 1998.

[3] N. Buchbinder, K. Jain, and J. S. Naor. Online primal-dual algorithms for maximizing ad-auctions revenue. In *Proceedings of the 15th Annual European Conference on Algorithms*, ESA'07, pages 253–264, Berlin, Heidelberg, 2007. Springer-Verlag.

[4] B. Chen. AT&T allows advertisers to sponsor mobile data. http://bits.blogs.nytimes.com/2014/01/06/att-allows-advertisers-to-sponsor-mobile-data, 2014.

[5] N. Economides and B. Hermalin. The economics of network neutrality. *RAND Journal of Economics*, 43(4):602–629, 2012.

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

[7] P. Hande, M. Chiang, A. Calderbank, and S. Rangan. Network pricing and rate allocation with content provider participation. In *INFOCOM*, pages 990–998, 2009.

[8] KDB_Daewoo_Securities. Telecom service 2014 key theme: LTE ecosystem. http://www.kdbdw.com/bbs/download/181382.pdf?attachmentId=181382, 2014.

[9] S. Lahaie, D. Pennock, A. Saberi, and R. Vohra. 'sponsored search auctions. In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory*. Cambridge University Press, Cambridge, UK, 2007.

[10] P. Maillé, E. Markakis, M. Naldi, G. D. Stamoulis, and B. Tuffin. Sponsored search auctions: An overview of research with emphasis on game theoretic aspects. 12(3):265–300, Sept. 2012.

[11] A. Mehta, A. Saberi, U. V. Vazirani, and V. V. Vazirani. Adwords and generalized online matching. *J. ACM*, 54(5), 2007.

[12] D. Mitra and Q. Wang. A model-based study of the impact of managed services and the spawning of aplications in broadband networks. In *Proceedings of the Workshop on Telecom Economics, Engineering and Policy, 24th International Teletraffic Congress (ITC 2012)*, 2012.

[13] P. Njoroge, A. Ozdaglar, N. Stier-Moses, and G. Weintraub. Investment in two sided markets and the net neutrality debate. Technical Report DRO-2010-05, Columbia Working Paper, 2010.

[14] Y. Wu, H. Kim, P. Hande, M. Chiang, and D. Tsang. Revenue sharing among isps in two-sided markets. In *INFOCOM*, pages 596–600, 2011.

[15] Y. Zhou, D. Chakrabarty, and R. M. Lukose. Budget constrained bidding in keyword auctions and online knapsack problems. In *Proceedings of the 3rd Conference on Web and Internet Economics*, WINE'08, 2008.