Reconciling Bitter Rivals: Towards Privacy-Aware and Bandwidth Efficient Mobile Ads Delivery Networks

Aruna Seneviratne*, Kanchana Thilakarathna*, Suranga Seneviratne*, Mohamed Ali Kaafar*‡ and Prasant Mohapatra†
*School of EE&T, UNSW and Networks Research Group, NICTA, Sydney, Australia
†Department of Computer Science, University of California, Davis, CA 95616
‡INRIA, France

***Abstract*—The use of free mobile services and applications (commonly referred as *apps*) are becoming increasingly popular. Such services and apps are generally monetized by means of third party advertising. The app developers and ad networks which provide the advertisements to be displayed within apps use every means to maximize their revenue, most often at the expense of the end user. These means of maximizing revenue impact the user in three ways: 1) Through the loss of privacy and control over their data; 2) through the increase in monetary cost due to communications overheads introduced by ad traffic; and 3) by the increase of battery usage. The introduction of rich media advertisements will have even greater implications with respect to the aforementioned bandwidth and battery consumption concerns. In this paper, we propose a novel architecture, called MASTAds, that combines the concepts of opportunistic networks, network intermediaries and predictive regularity of human behavior which enable both cost and energy-efficient ads delivery. In addition, MASTAds allows ad networks to obtain only the necessary information to provide targeted advertisements and high Ads revenues, whilst still preserving the user privacy.**

I. Introduction

The new mobile devices (e.g. smartphones and tablets) ecosystem is generally based on the use of applications or *apps*, which provide a vast range of services to their users. Typically, users download apps via applications markets. The apps available in these markets are either provided for free or for a fee. A recent study [1] has shown, that different online markets comprise different percentages of available free apps. For example, up to 70% of the apps are free in the Google play store, and 66% for Windows mobile apps centre, 42% for the Blackberry app centre and 47% for the Apple app store. Similar percentages have been reported by other studies as well, e.g. [2], which suggests that approximately 50% of the apps, across the major applications market places are free. Even though the fraction of free apps is already significant, we believe that this number will increase in the future as users will get used to the availability of more and more free apps.

The development of the free apps is made possible by means of the advertising revenues. App developers get paid by the *ad networks* for providing space within the app to display the advertisements and for collecting and providing user information which enable the ad networks to serve target advertisements [3].

In addition to the loss of privacy due to user information being passed on to the ad networks, the free app users incur other hidden costs. Firstly, the advertisement traffic contribute to the users' data downloads. With operators moving towards capped plans which meter data downloads, advertisement traffic can result in significant costs to the users. It has been shown that a popular mobile game can lead to as much as 40MB of extra data downloads per month [4]. More importantly, data downloads due to advertisements are likely to increase significantly when rich media ads, especially when video ads would become more prevalent. Secondly, the advertisement traffic can increase of the power usage of the mobile devices. For example, it has been shown that aggressive ad refresh rates cause the mobile devices to continuously be in high power states [5]. As a result, as much as 65% of the total energy usage of free apps can be attributed to advertisements.

Designing a system that preserves user privacy whilst minimising energy and bandwidth consumption is therefore a necessary, but challenging task. In essence, our aim is to design a new architecture that includes the following desirable features:

- Provide the necessary information to the ad networks to generate revenue. This would benefit to the app developers and the advertising agencies offering more targeted advertisements, i.e. maximizing *Usefulness*;
- Limit the loss of user privacy and offering users the control over their personal information, i.e. maximizing *Privacy*; and
- Minimize the resource usage both from the devices perspective (energy) and from the network resource consumption (bandwidth), i.e.maximizing *Efficiency*.

There has been a number of efforts which aim at developing mechanisms for protecting the *privacy* of users of online services [6], [7]. However, these works did not focus on mobile systems and more specifically on mobile applications, and as a result, they do not consider the system's *efficiency* as one of the constraints for the system design. In addition, there has been a considerable effort directed at understanding the system's

978-1-4673-5494-3/13/$31.00

efficiency of mobile systems, especially in terms of bandwidth and energy usage, when running free apps [4], [5]. These in contrast, do not address the issues of *privacy* loss. To the best of our knowledge, MobiAd in [8] is the only work that specifically addressed privacy issues in mobile advertising systems. However, the primary focus of MobAd is the preservation of privacy, and the system's efficiency is implicit as it proposes the use of GSM/3GPP broadcast channels for ad distribution and opportunistic networking for transmission of click reports for billing. Moreover, MobiAd only provides an overview of the architecture and unfortunately does not provide any details about the actual algorithms to be deployed.

In this paper, we explicitly address the issues of system *efficiency*, preservation of *privacy* without compromising the *usefulness* for ad supported mobile apps, by combining the concepts of decentralized personal data architectures [9], [10], [11] and delay tolerant networking [12]. In particular, we focus on applications that will be used for generating and distributing user generated content as we believe in the future more and more users will use their mobile devices to not only consume content, but to also create a vast amount of new content.

The rest of the paper is organized as follows. The next section highlights the problems of the existing advertisement distribution eco-system. In section III, we discuss the possible alternatives to determine the most suitable mechanism to address prevailing issues. In section IV, we propose a novel advertisement distribution architecture and present how it aligns with prevailing commercial ad networks. We discuss possible limitations of our architecture in section V. Finally, section VI presents the related work followed by the conclusion in section VII.

II. In-App Mobile Advertising and Implications

A. Operation: A simplified view

In the current mobile advertisement distribution architecture, mobile app developers make their applications via app markets such as Google Play Store and Apple App Store. Apps that support ads have space within the app to display advertisements. Advertisers place their advertisements with ad brokering companies, generally referred to ad networks in the likes of Google AdMob[1], Millennial Media[2] and InMobi[3]. App developers register their apps with an ad network, which enables the ad network to display advertisements in the space provided within the app. A successful rendering of an advertisement is referred to as an *impression*. This is done by the app developer by including an ad library provided by the ad network inside the app. When such an *ad supported* app is used, the ad network utilizes the user profile that it maintains, the advertiser's requirements and the price the advertisers are willing to pay to determine what ad to display.

Ad networks get revenue from advertisers for displaying (publishing) their advertisements. App developers get revenue from the ad networks, when an advertisement is published in their application or when an ad is *clicked on*. The more a user uses free apps, the more ad networks can collect information about the user, to update/build a profile. The profiling makes it possible for the ad network to display more targeted advertisements in the future.

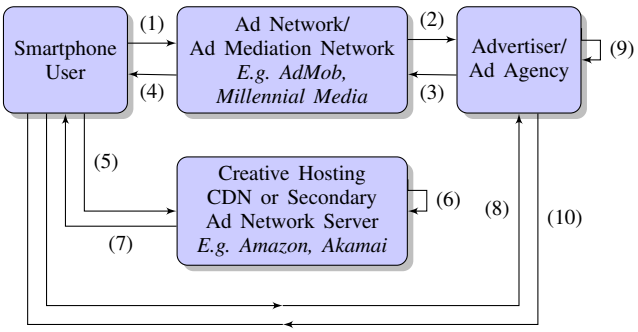

Fig. 1. Current advertisement distribution architecture

The ad distribution process is illustrated in Figure 1. The process starts when an ad supported app is used by the user. The app requests an advertisement from the ad network (step 1). Then the ad network selects an advertisement and provides the addresses of the related creatives (e.g. images/animations). Optionally, the ad network can use a *Real Time Bidding (RTB)* scheme to allow advertisers or agencies to dynamically bid for ads to be displayed. The ad network selects an advertisement using some criteria such as second price auction [13] (steps 2 and 3). Once the advertisement is chosen, the ad network sends the advertisement information such as the addresses of the creatives to the app (step 4). The app in turn initiates the fetching of the advertisement information (step 5). Creatives are generally hosted in *Content Distribution Networks (CDNs)* [5] or in different locations of the same ad network. One of *creative requests* is used to update the count of impressions as shown in step 6. Some advertisers run by their own count updates and analytics operations. In this case, there might be some additional steps, e.g. 8, 9 and 10 to fetch a *web beacon* (usually a 1x1 pixel) or a *302 URL* redirect [14].

B. Implications

The above advertisement distribution architecture has two major implications. Firstly, steps 1, 4, 5, 7, 8 and 10 can lead to significant additional data transfers and energy usage. Secondly, the collection and transferring of data to the ad networks can lead to loss of privacy of the users.

1) Additional Data Transfers: In [4], the example of the free game *Fruit Ninja* illustrates the additional data transfers and the corresponding cost of ads in the mobile environment. A 30 minute use of the free game *Fruit Ninja* per day, can consume up to 40MB of data per month. Hence, the use of 10 of such apps would result in a monthly 400 MB of data usage. For example, in Australia a typical 2GB monthly data plan costs around $50 and 400MB would then correspond to a cost of $10.

[1]http://www.google.com/ads/admob/
[2]http://www.millennialmedia.com/
[3]http://www.inmobi.com/

TABLE I
MOBILE RICH MEDIA AD EXAMPLES

Ad Network	Ad Type	Max Size (MB)
Yahoo [18]	Mobile Expandable	2.2
MobClix [19]	Full Screen Video	3
Millennial Media [20]	Interactive Video	0.3

On the other hand, recent reports show that the popularity of the rich media ads is increasing. For example, [15] shows a 23% increase in rich media ad impressions in Q2 compared to Q1, in 2012. In contrast, banner ad impressions decreased by 16% during the same period. Therefore, the extra data that ads generate is likely to grow significantly in the near future. Unlike mostly static banner ads which have simple dimensions and size restrictions (up to 15KB) [16], rich media ads come in various formats and implementations such as floating banners, expandable banners with video, interstitial interactives and pre-roll interactives. However, presently there are no specific standards for size, shape, function of mobile rich media ads [17]. Table I show examples which indicate what the sizes of potential future advertisements could be in the range of few MBs. In an attempt to unify various rich media ad formats and to create a common standard, IAB [21] has came up with Mobile Rich Media Ad Interface Definitions (MRAID). Most of the major ad networks are collaborating with IAB and it is expected that a common format will be adopted in the future. Nevertheless inclusion of videos and interactive parts means that the sizes of rich media ads will only exacerbate the problem, increasing the volume of data transfers of mobile users.

2) Energy: Current advertisement distribution process is found to be a major cause behind the smartphone battery drain. Pathak et al. [22] shows that 65%-70% of energy of free apps is "consumed" by third party advertisement libraries.

Qian et al. [23], [24] showed that, when using UMTS networks, most of the default ad refresh rates of popular ad networks cause the smartphones to be constantly in high power consuming states. Vallina-Rodriguez et al. [5] experimentally compared the smartphone energy consumption of three popular ad network libraries by means of a purpose-built Android app, for both WiFi and 3G networks. Their results further show that lower refresh rates yields to lower energy consumption when compared to the baseline consumption, for both types of networks.

Another aspect of advertisement affecting the smartphone performance is the recurrent download of static objects. By monitoring the 1000 most popular objects in advert traffic, authors in [5] also showed that up to 95% of the ads traffic can be redundant.

3) Privacy and Targeting: Recent work by Grace et al. [3] shows that popular ad libraries in free apps, collect other information such as location, other installed apps on the device, and bowser bookmarks. Some ad libraries even collect the IMEI number of the phone and permissions for other apps installed to receive remote commands. This not only a severe breach of users privacy, but it also compromises the integrity of the device as it can be used for illegal purposes. Similar work by [2] revealed that 7 out of 10 free apps in the Android market ask for dangerous access permissions such as access to SMS content and call history.

III. SYSTEM DESIGN ALTERNATIVES

An ideal system needs to have privacy mechanisms that enable the collection of the necessary information for targeted advertising, accounting purposes and fraud detection without the system being able to associate any information that is collected (e.g., clicked ads) with a particular user or any privately identifiable information. Moreover, the privacy preservation mechanisms that are used to achieve the above should not (a) limit the effectiveness of the auction mechanisms, and (b) negatively impact usage of system resources particularly, bandwidth and energy through introduction of extra advertisement-related data exchanges.

This section attempts to look at alternatives and determine the characteristics of the most suitable mechanisms for addressing each of the above aspects. Then it uses the findings to define the characteristics of a system that addresses the issues of resource usage, privacy and usefulness as a whole.

A. Privacy Preservation

As described earlier, the privacy preservation mechanisms need to provide the same usefulness, without compromising users privacy. This needs to ensure that the network level information, user activities and interactions with the network cannot be associated to identify a particular user.

In the case of network information, there have been several proposals. All of the proposed schemes adopt an approach which routes personal information through intermediaries [6], [8]. On the other hand, the proposed solutions for preventing leakage of privacy through user activities and interactions broadly fall in to one of the following categories:

- Keeping the data in a private data store on the client device and providing it to service providers as required in different forms and under different assumptions. For example, if it can be assumed that the service providers will not store the information, the data can be provided as required by the service providers [25]. Alternatively, computations or transformations can be allowed on the client device, using third party transformations and then only providing the results to the service providers [26], [27].
- Using a trusted/untrusted intermediary acting as a relay (proxy) [6]

The above suggests that using an intermediary which can be semi-trusted provides a good compromise for providing the necessary privacy preservation mechanisms as it has the capacity to provide both network level and user level privacy preservation mechanisms. This is also in line with the standardization activities such as *do not track* [28]. One of the key challenges then, is to develop a system architecture with intermediaries that does not limit the effectiveness of the

auction mechanisms and negatively impact the system resource usage.

B. Account Keeping and Fraud Detection

The system needs to enable the ad networks to charge the advertisers and pay the app developers/publishers for the advertisement that are displayed/clicked, i.e. account keeping. In addition, the system needs to detect the fraudulent behavior of the users, e.g. the use of bots for clicking on advertisements. The main challenge here is to extract the necessary information that enables correct account keeping and fraud detection, without compromising the privacy of the users.

For account keeping, in general cryptographic solutions, such as zero knowledge proofs, electronics tokens and mixing have been proposed [29], [7], [6]. These techniques have been shown to provide adequate solutions for both the *charge per click* as well as *charge per impression* models. In contrast, the detection of click-fraud is more challenging when privacy preserving techniques are used, simply because they hide some of the clients interactions from both the intermediaries and ad networks. So far, there has not been any approach that has been shown to effectively detect click fraud in privacy preserving systems.

This suggests that an adequate solution should leverage a cryptography-based approach for account keeping. Moreover, a number of click fraud detection mechanisms can operate in parallel to prevent or at least detect potential click frauds.

C. System resource usage

The optimization of system resources, namely bandwidth and battery usage has been addressed by a number of research papers.

1) Bandwidth: The schemes that have been proposed for minimizing bandwidth usage have adapted two generic traffic optimization techniques that exploit the specific characteristics of mobile advertisements. The first class of approaches adapts caching techniques, as the mobile advertising traffic has been reported to have significant amount of redundancy [30]. In addition, the delay tolerant nature of mobile advertisements and the fact that advertisements consists of a collection of static components, caching could lead to significant bandwidth savings. The second class of approaches exploits the predictability of users to intelligently schedule advertisement traffic to maximize the use of low cost WiFi networks [4]. Other mechanisms use transport channel characteristics such as the availability of broadcast channels [8]. Since broadcasting can avoid multiple dedicated downloads to separate users, it can help to reduce bandwidth usage.

Of these techniques, the use of transport channel characteristics may prove impractical or have limited deployment possibilities, as they may not be universally available. In the case of caching, it is necessary to consider implications of the use of *cache busting* techniques to obtain better control by some of the players, e.g. ad networks. Furthermore, caching mechanisms may also have implications on other mechanisms of the advertisement system such as the real time bidding. This suggests that the generalized bandwidth optimization mechanisms, which take into account the specific needs and exploit the characteristics of the advertisement traffic are best suited for optimizing the bandwidth usage.

2) Energy: The work related to energy-savings which focus on advertisement systems, fall into two broad categories: those that use of traffic shaping and those that use of networks that have lower transmission power requirements.

- Traffic shaping: These methods focus on the UMTS networks where the mobile devices operate in high power, low power and idle states. They attempt to minimize the time a mobile device is in the high power state, and the number of state transitions [23].
- Low transmission power: WiFi networks have lower energy per transmitted bit compared to UMTS networks [31]. Therefore, these schemes attempt to minimize the energy usage by using WiFi networks though the use of intelligent scheduling.

Both these categories provide effective solutions. Their use will depend on the system architecture and the type of networks that are being used.

IV. System Architecture

We propose a new mobile advertisement distribution architecture, called MASTAds (Mobile Anonymous but Still Targeted Ads), that reconciles the conflicting constraints of privacy, energy and bandwidth consumption while still enabling the delivery of targeted advertisements. It uses an intermediary, namely an *Advert Management Server (AMS)*, between the users and the ad networks as illustrated in Figure 2, which can be easily integrated to the current ad eco-system. We assume that the AMS acts as a semi-trusted broker for the users and collects information about users' interests anonymously and distributes targeted advertisements to users. In practice, the AMS can be hosted by the mobile network operator or within an ISP network (or at a higher level) depending on the velel of privacyy protection that need to be provided. Generally, the main reason behind introducing an intermediary in between the users and a central entity such as an ad network entity, is to hide users' sensitive information from the ad network as described in section III-A. Since the system design goals of MASTAds is also to minimize resource consumption, the AMS constructs communities that would allow a robust and resource-efficient delivery of the advertisements, as described below.

A. General overview

The MASTAds architecture uses the users' contact patterns to build highly connected and robust communities of users. This contact-based community formation is done prior to and independently of the advertisement delivery processes. The community creation process is similar to the creation of tribes in [32] and is described further in section IV-B. Once the communities are established, users belonging to a community can indicate their interests to the AMS. These interests can be expressed using advertisement categories and can either

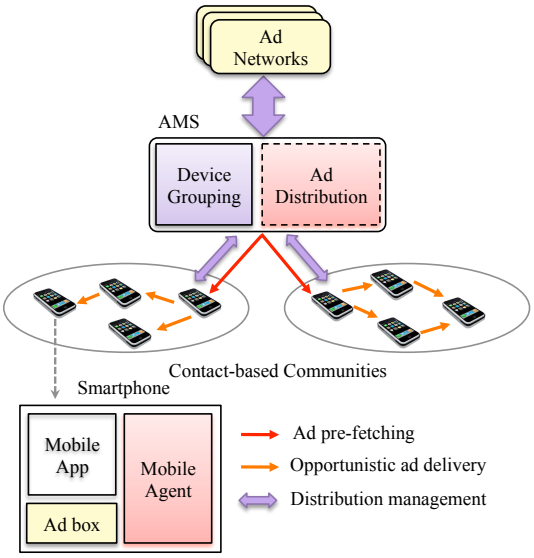

Fig. 2. System architecture

be broad or fine grained, as discussed in section IV-C. The interests of the whole community is then used by the AMS to obtain advertisements from the ad networks and distribute to the users.

The process of users' interest indication is designed so as not to reveal the interests of an individual user to other users within the community or the AMS. This process is also described in section IV-C. In this paper, we focus on what is commonly referred to in-apps ads (ads embedded into mobile phones applications) and as such we can envision several ways of extracting and representing users interests. First, users can explicitly express interests in different ads categories through the an API provided to the applications developers by the ad networks. The interests can also be derived from the data of the other applications, with appropriate permissions. Finally, the interests can be represented by the categories of applications in which advertisements would be displayed.

Once the AMS obtains the relevant advertisements that corresponds to the interest of a community from the ad network, they are pushed to the relevant community together with the expressed interests. Again, the *ads delivery* process is designed to ensure that users' interests or targeted advertisements cannot be inferred. This process is described in section IV-C3 and section IV-D.

MASTAds minimizes bandwidth and battery usage by adopting a pragmatic multi-processes scheme. Once the communities are identified by the AMS, it selects some of the devices, generally the devices with the highest degree of connectivity, within the communities as initial propagators of both ads and interests. Then, it uses the propagators for interest propagation to the AMS and ad delivery from the AMS to the users as described in section IV-C .

Once a user device gets the advertisements relevant to its own interest categories, a *Mobile Agent* within the device handles the *ads displaying* process according to the advertisement attributes and displaying policy as detailed in section IV-E.

B. Contact-based Community Establishment

Each device maintains an individual contact graph and updates the AMS periodically. AMS uses individual contact graphs to detect contact-based communities, i.e. it creates an aggregated dynamic contact graph. Due to the regular mobility patterns and connectivity among users, it is expected that the communities converge to a steady contact graph over time, as described in [33].

When a new device joins the system, it is assigned to an initial community. The choice of the initial community will depend on the devices it had encountered. The contact graph is then refined and the AMS determines the community to which the new comer will be assigned. This process does not require the devices to reveal their locations. They only need to reveal their contact graph with the other devices over time.

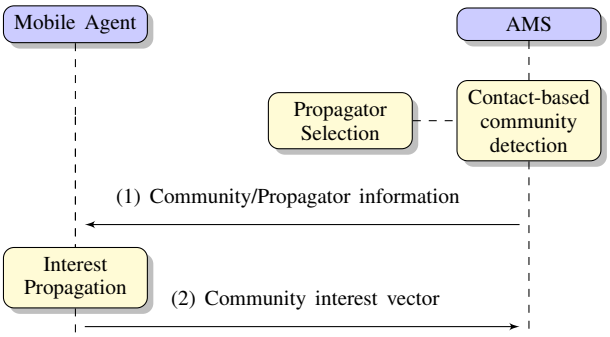

Fig. 3. Contact-based community detection and interest propagation phases of MASTAds

We leverage opportunistic contacts between the different devices in the communities. Therefore in MASTAds a set of propagators are selected prior to the interest propagation and advertisement delivery phases. This results in bandwidth savings for devices within each community as the data transfers are done opportunistically and locally.

The advertisements delivery delays and success rates will thus highly depend on the size and the behavior of the set of propagators. Hence, we aim to take advantage of routine behavioral patterns of mobile users for the propagator selection. Typically, users have daily (resp. weekly) routines and with a high likelihood that each user's device will be in contact with a returning set of devices every day (resp. week). Specifically, as shown in Figure 4, the AMS can select propagators for the week $k+1$ based on aggregated mobility patterns as observed up to week k. We virtually divide every week into Δ time slots, where Δ represents the delivery deadline for a set of advertisements. Then, the AMS selects a set of propagators P_i for each Δ_i for each community. The propagator selection algorithm has to ensure fairness in resource usage among the community members, and as such there is a trade-off to consider between the ads delivery performance in terms of propagation delay and success rate and the size of the set of propagators. Thus, the primary objective of the propagator selection is to identify

the minimum cardinality set of propagators which satisfies the delivery requirements such as the delivery success rates and delivery deadlines. This is equivalent to the helper selection problem presented in [33]. Hence, the greedy helper selection algorithm can be easily adopted to select the propagators in MASTAds.

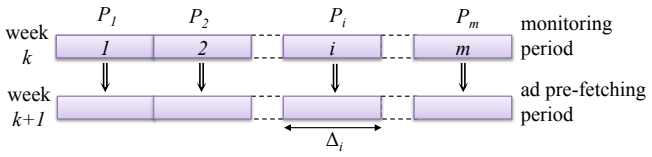

Fig. 4. Periodic weekly propagator selection

C. Interest Propagation and Advertisement Pre-fetching

In contrast to the current advertisement distribution architecture, we advocate that the user profiling for the purpose of privacy-preserving targeted advertisement delivery be carried out on the user devices. We envision several ways the devices-level profiling can be handled. Mobile agent can continuously monitor background activities such as social networking behavior, mobility and connectivity patterns of the user and create an interest-based user profile. Users can also express their own interests in advertisement categories, similarly to the current advertisement preferences system provided by Google[4] and opt-in a specific advertisement preference system hosted at the level of their device. The interests can also be extracted from different combinations of permissions of the applications installed on the device has gathered.

1) Privacy-preserving interests propagation: It is vital that the interests collected at the level of the device in the MASTAds architecture is protected to ensure privacy. In essence, no other entity should be able to distinguish particular interests sent by one specific device. The other users in the community should not be able to learn the specific interests of any initiator of advertisement propagation. Similarly the selected propagator should not be capable of distinguishing who initiates an interest propagation phase, and the devices that collected interests. The AMS may learn accumulated (aggregated) set of interests belonging to the community the propagator is representing, but should not be able to distinguish individual interests. Finally, the ad network does not need to know the particular interests of a user, but simply needs to know that delivered ads are highly targeted. This can be achieved by allowing the ad network to maintain an aggregated view of the interests expressed by the community, through the interests collected and transmitted by the AMS. This is achieved using a probabilistic interest dissemination scheme that operates on top of the identified community links as described below.

2) A probabilistic interests propagation phase: Let's consider a user A with interests (or part interests) I_{a1}, I_{a2} that have to be sent to the AMS. For the sake of simplicity, let's now assume that A has already received from other users B and C, their interests $I_{b1}, I_{b2}, I_{c1}, I_{c2}$. A can then concatenate interests and forward an interest vector to other devices that it is in contact with a configurable probability, p_f. The portion of self-interest categories added to each interests vector can also be determined according to the level of privacy the user intends to maintain. A then decides on which links the interests vectors are to be forwarded. As a result, a forwarding of an interest vector does not enable the identification of the source of interests. Colluding users might compromise the anonymity of such a probabilistic interest dissemination process, since many colluding users observing whether an interest has been transmitted or not may suggest whether or not the user being observed is effectively the source of the interests. Nevertheless, this potential attack requires the colluders to be directly connected to the targeted user. In addition, as described later, MASTAds addresses the potential colluders attack and the interests initiator privacy issues by transmitting a dummy set of interests with each set of initial interests transmitted to the community. This ensures that even when an interest vector is tracked back to its originator, the attacker cannot distinguish between the actual interests of the initiator and the dummy interests.

As illustrated in Figure 5 each device keeps track of the link to which the interest vector (whether accumulated interests or interests padded with dummy interests, I_1) have been delivered. Whenever it is possible, the link used to propagate the interests will also be used to receive the corresponding advertisements. This ensures privacy during the advertisement delivery process as described in section IV-D.

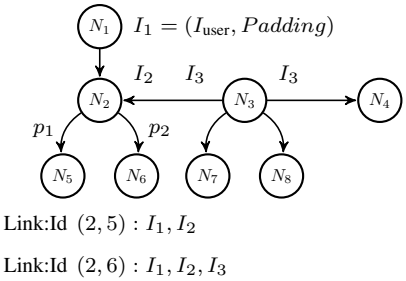

Fig. 5. Probabilistic interest propagation

Let's now consider the case of a new user joining a community. We examine the case where the new user's interests can be identified from the existing users' interests (e.g. all the users actual interests are different from the existing community members interests). If the AMS keeps track of each community interests, with a new user joining the community the AMS might be capable of building the new user's profile by inspecting the difference in terms of previous and current community interests. Similarly, a curious community member can infer a new user's interests when this newcomer forwards interest messages to which it has added its own interests. Again, including dummy interests as well as the user's actual interests in the interest vector can solve this. The use of dummy interests has to also take into account the cases where users leave a community. If a users interests has very

[4]See www.google.com/ads/preferences

little overlap with other community members, the lack of information occurring after the user leaves a community can allow an adversary to infer the user interests. In MASTAds this is overcome using a minimum number of dummy interests in each interest vector. This results in a minimum level of anonymity among the users interests. E.g. k categories chosen at random being added to each users' interests I_i. Even though this will introduce an overhead, it will in turn guarantee at worst a k-anonymity level for every user that joins the system.

3) Advertisement pre-fetching: When the accumulated interests vector reaches one of the selected propagators, it will forward the aggregated "community interest vector", I to the AMS. Typically, the selected propagators send an advertisement request to the AMS along with I, when they are connected to a low-cost network. AMS requests advertisements from the ad networks based on a few community attributes such as the community interest vector, the community size and location (step 2 in Figure 3). Ad networks auction the attributes of the communities, thus allowing advertisers or ad agencies to bid for the ad slots (step 3 and 4 in Figure 6). The, ad network then selects and delivers advertisements to AMS according to the ad selection policy among advertisers and ad networks (step 5 in Figure 6). The advertisement selection policy attributes are embedded into the advertisements since it has to be reconsidered at the mobile agent when displaying the advertisements. After receiving advertisements from an ad network or a set of ad networks, the AMS bundles the advertisements together based on advertisement attributes and distributes the bundled advertisements to the selected initial propagators of each community. Mobile agents on the user device use this to maintain a cache of pre-fetched advertisments: add new advertisements to the cache and remove expired advertisements from the cache.

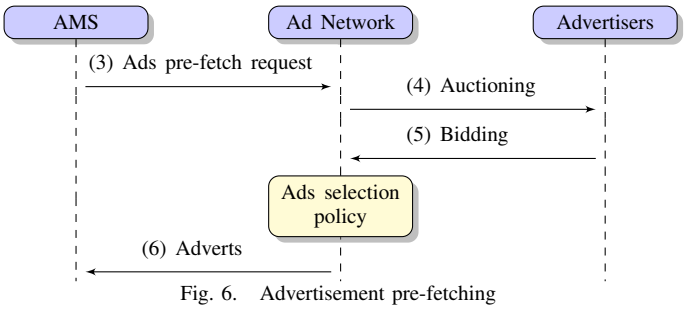

Fig. 6. Advertisement pre-fetching

D. Advertisement Delivery

In MASTAds, the AMS entity expects the propagators to disseminate the advertisement within the community through opportunistic direct communication (step 7 in Figure 7). When two devices are connected to the same WiFi access point, we consider that these two devices are in direct communication range of each other. The interests propagators selection ensures that advertisement distribution satisfies delivery delay constraints and provides the required success rates[5].

[5]In this distribution process a peer to peer protocol such as trackerless BitTorrent can be used to increase the scalability.

Similarly to the interests propagation phase, advertisement distribution process has to ensure a level of anonymity so that other entities participating in the delivery process cannot link the targeted advertisement to the device for which the advertisement are intended. In other words, the advertisement delivery process should not leak information about users interests. As demonstrated in [34], advertisement eavesdropping can reveal a significant fraction of a users online activities (e.g. browsing history). In MASTAds this is avoided as follows. Since each device maintains a record of the path that interest vectors were propagated during the interest propagation phase, the advertisement delivery phase has to give higher priority to distribution of advertisement along the links used for the corresponding interests propagation as described in section IV-C2. Namely, whenever the link that has been recorded as carrying the corresponding interest is available, the device chooses it to deliver the advertisement relevant to such interests.

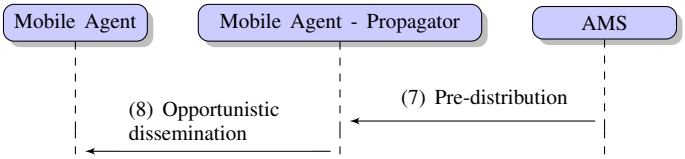

Fig. 7. Opportunistic advertisement delivery

E. Advertisement Displaying

Once the advertisement pre-distribution has been completed, every device will have a set of cached advertisements. When the user launches an ad supported mobile app, advertisement request is directed to mobile agent instead of the ad network as shown in Figure 8. Each advertisement comes with a set of attributes such as a priority value, preferred location and mobile apps. The priority value will be related to the delivery deadline of the advertisement. This is determined by the contract between ad network and advertiser. Then, the mobile agent delivers suitable advertisements based on advertisement attributes, advertisement display policy and real time contextual information such as location of the device and time of the day enabling fine grained targeted advertising.

Since AMS pre-distributes a set of advertisement, at each advertisement refresh request mobile agent delivers the same or a different advertisement based on the ad display policy. Therefore, the mobile device is not required to move to a high power state to fetch advertisements from the ad network. Thus, the energy consumption of the device for advertising can be significantly reduced. Also many mobile apps only require internet connectivity for advertising related activities. These apps, currently allow the users to disable internet access and continue to use the apps without advertisements. This results in a revenue loss for the app developer and the ad network. MASTAds in contrast allows displaying advertisement even when the app user does not have internet access. potentially resulting in more revenue opportunities for ad networks and app developers. It also enables, mobile apps to accommodate higher advertisement refresh rates for higher impression

counts, which again increases the revenue for both ad networks and app developers.

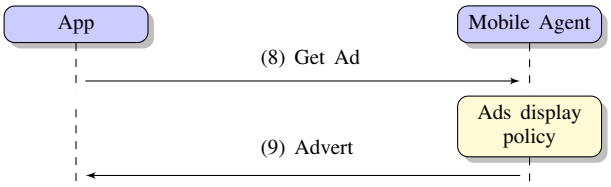

Fig. 8. Advertisement displaying

F. Ads Impressions and Clicks

In MASTAds each mobile agent keeps records of the number of advertisement impressions and anonymously report the number of impressions back to the ad network. It uses several techniques to minimize the resource consumption whilst preserving user privacy. First, impression reports are encrypted with the ad network's public key so that the AMS cannot trace back advertisement impressions to a device. Secondly, to hide the user's identity from the ad network, it uses anonymous routing (e.g. Tor), similarly to the MobiAd [8]. While an anonymizing network might introduce relatively high delays, we stress that based on our observations indicate that ad networks do not report impression counts in real time. For example, Google AdMob mentions that clicks and impression reports can be delayed up to 48 hours. Therefore opportunistic anonymous distribution of these reports does not impact the system performance. However, it results in both bandwidth and energy savings.

In the current mobile ad distribution architectures, when a user clicks on an advertisement, it first contacts the ad network (reporting the click), which redirects the user to the advertisers landing page. As a result, both ad network and the advertiser are aware of the click source and are able to collect information about the users interests. In MASTAds, we separate the two different processes as they have different requirements.

An ad click redirects the user to the advertiser landing page and needs to be handled in real time. In such a case, where the user is not willing to exposes interests to both the ad network and the advertiser, MASTAds will use an anonymizing network such as Tor to visit the landing page. For billing purposes, both the AMS and the ad network need to know whenever there is a click on an ad. For these exchanges, similarly to the ad impression process, MASTADs uses anonymizing opportunistic routing.As the click reports are delay tolerant this can be used without impacting the system performance.

V. Discussions and Possible Limitations

MASTAds is a novel architecture that aims to reconcile both privacy and resource consumption in mobile advertisements delivery networks. We leverage opportunistic and probabilistic routing along with a device-centric expression of interests to reach a balance between bandwidth and energy consumption while preserving users privacy. In essence, the aggregation of users interests as well as advertisement reports within communities, now provides a view at the community level to the ad network as opposed to uniquely trackable identity of users, ad clicks and interests. This aggregation of users interests will impact some of the profiling techniques that are currently implemented based on individual operations by major actors in the advertisements delivery networks. This needs to be carefully studied.

In particular, as MASTAds hides the source of click and impression reports, the current click fraud prevention methods might not be directly applicable. An ad network however can still detect a click fraud based on detection of anomalous deviations from regular ad impressions and clicks patterns at a community level rather than at a single IP address level. Identifying which member of the community is potentially triggering such a deviation might however be challenging. In fact, the community building process is hidden from the ad network. Furthermore, even though the AMS controls the members of each community, the source of click and impression reports are unknown to the AMS. As a result, MASTAds architecture does not provide sufficient information to the AMS to identify aggressive ad clickers. One possible solution is to rely on the community members themselves, which can identify abnormal activities of reports, and then report to the AMS some observed statistics on different links they maintain. These statistics gathered at the level of the AMS can be a good indicator of suspicious clicks volumes on links, and hence identify their originators. Unfortunately, due to the opportunistic nature of reports dissemination, a malicious device can still try to dilute the reports in time so that the frequency of reports sent over each link seems innocuous. An efficient click fraud prevention mechanism on top of MASTAds needs further study and is left as an open research issue.

MASTAds uses device to device opportunistic communication in a probabilistic way to propagate interests and advertisements. While the number of transmissions may appear to be redundant compared to a stand alone advertisement download, several factors contribute to the bandwidth and energy efficiency of the system. User interests are generally stable for a period of time. As a result their interest vector would only change very slowly. Thus devices will not propagate new interests regularly. Advertisement dissemination frequency can be a critical factor causing high bandwidth and energy cost. However MASTAds opportunistically use low cost networks such as WiFi that enables the use of lower cost network resources. As a result the overall resource efficiency of the system depends on the availability of low cost networks and the validity periods of the cached advertisements. We believe ad pre-fetching and delivery strategies can be dimensioned considering such factors but needs further investigation.

The pre-fetching process of advertisements requires the ad networks to decide the advertisements to be pushed in advance and hence the current real time bidding systems need some revision. For example the ad network can auction the total slots in the community, giving an indication on the maximum

delay in displaying. If energy constraint can be relaxed the mobile agent can anonymously participate in the real time bidding process similar what was proposed in [6] via the AMS. Therefore an energy efficient real time bidding scheme for MASTAds needs to be studied further.

VI. Related Work

There have been numerous alternative proposals for advertisement delivery. However, a majority of them donut satisfy least one the three requirements: *usefulness*, *privacy* and *efficiency* . In mobile networks all these aspects are equally important and here we discuss some of the recently proposed ad delivery architectures in this context.

Privad [6] and Adnostic [7] are two alternatives that attempt to preserve users privacy in online advertising. Privad does not reveal any user information to ad networks. However it is less practical in terms of deployability, compared to easily deployable *cookies* based server tracking[6]. On the other hand Adnostic lies in between these two extremes: cookies-based tracking and Privad. In order to provide the anonymity and hide user interest, both methods require the fetching of large amount of extra advertisements to the user device. In Privad, an average monthly download around 30MB is required to provide a reasonable pre-fetching of search sponsored advertisements, each sized of 250B. However a typical mobile banner ad is around 10KB and as a result monthly bandwidth consumption will then be around 1.2GB. Thus, such methods cannot be effectively used in mobile networks due to their relatively intensive bandwidth and energy consumption.

MobiAd [8] is the only work that specifically addresses privacy in mobile ad delivery. MobiAd disseminates location specific targeted ads to mobile users by broadcasting the advertisements via local base stations or access points and the relevant advertisement are cached on the devices. As a result, there is no cost to the users. Authors have proposed to use DTN to transfer the encrypted information about impressions and clicks, so that intermediate nodes and advertisement brokers cannot determine which user viewed which advertisement. However, the proposed broadcasting technology (Multimedia Broadcast and Multicast Services-MBMS) for UMTS supports only small text based ads. Moreover, the effectiveness of broadcasting of advertisements largely depends on the size of the advertisement inventory and is suitable only for small advertisements.

On the other hand, there have been proposals to address resource consumption in mobile advertising. Khan et al. [4] proposed CAMEO [Context Driven Advertisement Modulator and Optimizer] which is also propose the use of a local cache similar to Privad [6]. It differs from Privad, as it tries to infer the future context of the device so that it can be used as an input to the advertisement pre-fetching process. Moreover CAMEO will always try to use low cost networks such as free WiFi for advertisement downloading so that no costs are associated with the pre-fetching process. CAMEO also addresses the bandwidth and performance issues associated with ad distribution. However the architecture will allow the ad networks to store all the contextual information about the user, thus does not consider user privacy.

Adcache [5] is a mobile advertising solution, in which an agent at the mobile device pre-fetches advertisements and maintains a cache of advertisements in order to save bandwidth and energy. To preserve privacy, similar to what was proposed by Grace et al. [3] and Leontiadis et al. [2], permissions required to provide advertisements are decoupled from the permissions required for the core operations of apps. However limited information is given about the ad pre-fetching process which ultimately will decide the quality of targeting. If the solution does not support accurate targeting, the usefulness of the scheme is limited as there is no real incentive for ad networks to adopt it.

[6]http://33bits.org/2012/06/11/tracking-not-required-behavioral-targeting/

VII. Conclusion

We studied the existing in-app mobile advertising ecosystem and identified characteristics and features that can be effectively exploited to improve the system performance by minimizing the communications cost and energy usage. In addition, at the same time maintaining the privacy of the users and enabling the service providers to obtain the information necessary to continue providing the services effectively. We then used the insights gained to develop a novel advertisement distribution architecture, MASTAds, that satisfies constraints of minimizing system resource usage, maximizing usefulness to service providers and enhancing privacy of users. We illustrated how these different, often conflicting requirements, can be met by the proposed architecture that combines advertisement pre-distribution to devices that are connected to low-cost networks and using these devices to disseminate the advertisements locally to other users through opportunistic direct communication.

Although the concepts of using opportunistic networking and pre-distribution have been discussed in numerous application scenarios, to the best of our knowledge, they have not been used as proposed in MASTAds for optimizing systems along the three conflicting dimensions of resource usage, usefulness and privacy, especially in an advertisement distribution system.

MASTAds preserves user privacy without comprising the usefulness or the system efficiency. We anticipate the need of such a system in coming years will become vital as a result of the rapid popularity in mobile rich media ads along with privacy, bandwidth and energy concerns of mobile users.

References

[1] N. d'Heureuse, F. Huici, M. Arumaithurai, M. Ahmed, K. Papagiannki, and S. Niccolini, "What app? A wide-scale mesurement study of smart phone marketside-scale mesurement study of smart phone markets," *Mobile Computer Communication Review*, vol. Vol. 16, no. 2, pp. 16–27, April 2012.

[2] I. Leontiadis, C. Efstratiou, M. Picone, and C. Mascolo, "Don't kill my ads! balancing privacy in an ad-supported mobile application market," in *HotMobile 2012 - 13th Workshop on Mobile Computing Systems and Applications*, 2012.

[3] M. Grace, W. Zhou, X. Jiang, and A. . Sadeghi, "Unsafe exposure analysis of mobile in-app advertisements," in *WiSec'12 - Proceedings of the 5th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2012, pp. 101–112.

[4] A. J. Khan, V. Subbaraju, A. Misra, and S. Seshan, "Mitigating the true cost of advertisement-supported "free" mobile applications," in *HotMobile 2012 - 13th Workshop on Mobile Computing Systems and Applications*, 2012.

[5] N. Vallina-Rodriguez, J. Shah, A. Finamore, Y. Grunenberger, H. Haddadi, K. Papagiannaki, and J. Crowcroft, "Breaking for Commercials: Characterizing Mobile Advertising," in *Proceedings of IMC 2012*, 2012.

[6] S. Guha, B. Cheng, and P. Francis, "Privad: Practical Privacy in Online Advertising," in *Proceedings of the 8th Symposium on Networked Systems Design and Implementation (NSDI)*, Boston, MA, Mar 2011.

[7] V. Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas, "Adnostic: Privacy preserving targeted advertising," in *17th Network and Distributed System Security Symposium*, 2010.

[8] H. Haddadi, P. Hui, and I. Brown, "Mobiad: Private and scalable mobile advertising," in *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, 2010, pp. 33–38.

[9] http://joindiaspora.org.

[10] L. Cutillo, R. Molva, and T. Strufe, "Safebook: A privacy-preserving online social network leveraging on real-life trust," *Communications Magazine, IEEE*, vol. 47, no. 12, pp. 94–101, 2009.

[11] A. Mahdian, J. Black, R. Han, and S. Mishra, "Myzone: A next-generation online social network," in *Tech Report: Department of Computer Science, University of Colorado at Boulder*, 2011.

[12] K. Fall, "A dealy tolerant network architecture for challenged internets," in *ACM SIGCOMM*, New York, USA, 2003.

[13] B. Edelman, M. Ostrovsky, and M. Schwarz, "Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords," National Bureau of Economic Research, Tech. Rep., 2005.

[14] "Interactive Advertising Bureau: Mobile Web Advertising Measurement Guidelines," http://www.iab.net/media/file/MobileWebMeasurementGuidelines_final.pdf, 2011.

[15] "Opera software: The state of mobile advertising," in *http://www.opera.com/sma/2012/q2/*, Q2 2012.

[16] "Mobile Marketing Association: Mobile Advertising Guidelines Version 5.0," http://mmaglobal.com/mobileadvertising.pdf, 2011.

[17] "Mobile Marketing Association: Rich Media Mobile Advertising Guidelines," http://mmaglobal.com/rmma.pdf, 2011.

[18] "Yahoo! As Specs United States: Mobile Rich Media Ads," http://adspecs.yahoo.com/formats.php?id=55, 2012.

[19] MobClix, "Mobile Rich Media Ad Specs," http://www.mobclix.com/richmedia/pdf/adspecs.pdf, 2012.

[20] "Millennial Media," http://www.millennialmedia.com/files/resources/ad-specifications_10912.pdf, 2011.

[21] "Interactive Advertising Bureau: Mobile Rich Media Ad Interface Definitions (MRAID)," http://www.iab.net/media/file/IAB_MRAID_v2_FINAL.pdf, 2011.

[22] A. Pathak, Y. C. Hu, and M. Zhang, "Where is the energy spent inside my app?: Fine grained energy accounting on smartphones with eprof," in *EuroSys'12 - Proceedings of the EuroSys 2012 Conference*, 2012, pp. 29–42.

[23] F. Qian, Z. Wang, Y. Gao, J. Huang, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "Periodic transfers in mobile applications: Network-wide origin, impact, and optimization," in *WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web*, 2012, pp. 51–60.

[24] F. Qian, Z. Wang, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Profiling resource usage for mobile applications: A cross-layer approach," in *MobiSys'11 in Proceedings of the 9th International Conference on Mobile Systems, Applications and Services and Co-located Workshops*, 2011, pp. 321–334.

[25] M. Bilenko, M. Richardson, and J. Tsai, "Targeted, not tracked: Client-side solutions for privacy-friendly behavioral advertising." TPRC, 2011.

[26] M. Fredrikson and B. Livshits, "Repriv: Re-imagining content personalization and in-browser privacy," in *Security and Privacy (SP), 2011 IEEE Symposium on*. IEEE, 2011, pp. 131–146.

[27] R. Mortier, C. Greenhalgh, D. McAuley, A. Spence, A. Madhavapeddy, J. Crowcroft, and S. Hand, "The personal container, or your life in bits," *Digital Futures' 10*, pp. 11–12, 2010.

[28] J. Mayer, A. Narayanan, and S. Stamm, "Do not track: A universal third-party web tracking opt out," IETF draft-mayer-do-not-track-00, March 2011.

[29] M. Backes, A. Kate, M. Maffei, and K. Pecina, "Obliviad: Provably secure and practical online behavioral advertising," in *Security and Privacy (SP), 2012 IEEE Symposium on*. IEEE, 2012, pp. 257–271.

[30] F. Qian, K. S. Quah, J. Huang, J. Erman, A. Gerber, Z. Mao, S. Sen, and O. Spatscheck, "Web caching on smartphones: Ideal vs. reality," in *MobiSys'12 - Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, 2012, pp. 127–140.

[31] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: a measurement study and implications for network applications," in *Proc. of the 9th ACM SIGCOMM IMC '09*, Chicago, 2009, pp. 280–293.

[32] K. Thilakarathna, H. Petander, J. Mestre, and A. Seneviratne, "Enabling mobile distributed social networking on smartphones," in *Proceedings of the 15th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, Cyprus, Oct. 2012, pp. 357–366.

[33] K. Thilakarathna, A. C. Viana, A. Seneviratne, and H. Petander, "The Power of Hood Friendship for Opportunistic Content Dissemination in Mobile Social Networks," INRIA, Saclay, France, Tech. Rep., 2012.

[34] M. dung Tran, M. A. Kaafar, and C. Castelluccia, "Betrayed by your ads! reconstructing user profiles from targeted ads," in *The 12th (PETS 2012) Privacy Enhancing Technologies Symposium*, Vigo, Spain, 2012.