RI 12015 April 19, 2012 Computer Science

IBM Research Report

Exploiting Rich Telecom Data for Increased Monetization of Telecom Application Stores

Vikas Agarwal, Sumit Mittal, Sougata Mukherjea
IBM Research Division
IBM Research - India
New Delhi - 110070.

Palak Dalal
Indian Institute of Technology - Bombay
Mumbai, India.

Exploiting Rich Telecom Data for Increased Monetization of Telecom Application Stores

Vikas Agarwal, Sumit Mittal, Sougata Mukherjea
IBM Research - India
New Delhi, India - 110 070
Email: avikas, sumittal, smukherj@in.ibm.com

Palak Dalal
Indian Institute of Technology - Bombay
Mumbai, India - 400076
Email: palak@cse.iitb.ac.in

***Abstract*—With rapid advancement in the capabilities of mobile phones, there has been a tremendous increase in the development and use of mobile applications in recent times. Surprisingly, Telecom operators have not been prominent players in the success of mobile applications and the corresponding mobile Application Stores. We argue that, to offset this going forward, Telecom operators can utilize some of their core strengths to create differentiated Application Stores that will help them compete effectively with their competitors. More specifically, Telecom operators have access to unique data which is not available elsewhere; this can be used to give new insights on their customers and offer them relevant mobile applications. In this paper, we present techniques on how applications can be promoted effectively from a Telecom Application Store utilizing the users' location, their Telecom profiles as well as their social relationships derived from their calling patterns. Moreover, we provide a bidding platform that gives control to developers for highlighting their application by placing bids on multiple Telecom parameters. Simulations conducted utilizing real data from a large Telecom operator suggest the efficacy of our techniques.**

Keywords - Telecom, Application Stores, Promotions, Bidding

I. Introduction

With rapid enhancements in computing power, memory, display, etc., mobile phones have emerged as a platform for deploying and executing a plethora of applications. With the increasing popularity of mobile applications, 'Application Stores' offered by companies like Apple and Google have been hugely popular in recent times. For example, as of early July 2011, 200 million iOS users have downloaded over 15 billion applications from the Apple App-Store[1].

People access these applications from their mobile phones, but surprisingly, Telecom operators who provide the connectivity on these phones have been fringe players in the success of mobile applications. This is because the consumers are currently using the Telecom network as a mere dumb pipe to access compelling applications provided by mobile device manufacturers or Internet content providers. However, Telecom operators have access to some important data sources which are not available to their competitors like mobile device manufacturers and Internet content providers. We believe that if the operators can utilize their core strengths effectively, they can become primary players in the significant opportunity around mobile applications and Application Stores.

One problem with the popularity of Application Stores is that with the plethora of applications available, it is difficult for users to find out about applications that are relevant to them. Currently, Application Stores give top downloaded/rated applications, but this may not always be relevant to a user. Similarly, it is not easy for the store to determine which applications should be promoted to which users. The Telecom operators are in a better position to address this problem. In essence, they can use various unique data sources to promote the right applications to their consumers. For example, if an application is suited for a particular tourist spot, it can be recommended to users visiting that spot utilizing location information of users available to a Telecom operator. Likewise, a video streaming application can be attractive to users with an unlimited high-speed data plan. Moreover, Telecom operators can determine the social relationships of their consumers based on their calling and messaging patterns and recommend to consumers applications which were of interest to their friends.

To enable developers to promote their applications aggressively to the users, the Telecom operators can allow bids by developers on various Telecom parameters, such as specific locations and application categories like travel, social etc. For example, if a user is currently in Delhi, applications by developers who bid on location "Delhi" can be promoted. Likewise, for socially active users, applications by developers who bid on keyword "social app" can be promoted. Not only does this provide better control to developers towards highlighting their applications, it helps the Telecom operators generate additional revenues. Existing search engines, such as Google, allow advertisers to bid on keywords. However, a bidding platform that enables bids on 'multiple', 'complex' Telecom parameters does not exist. For example, when bidding on 'location', one needs to specify, apart from the bid value, information about geographical boundaries – country, city, or specific region(s) within a city – that the advertiser is interested in.

In this paper, we take a comprehensive view of various available Telecom data sources and analyze their utility in efficiently promoting applications to users in a Telecom-hosted Application Store. We believe that adoption of our techniques would help a Telecom operator to better monetize

[1]http://en.wikipedia.org/wiki/App_Store

it's offerings. In particular, our main contributions in this paper are -

- We present an architecture of a system called *TappStore* that utilizes various Telecom data sources not only for enabling efficient promotion of applications to users, but also for offering a bidding platform to developers.
- We describe promotion and campaign techniques that work on the notion of *relevance matrices* and associated *metadata tables* created through analysis of Telecom data.
- We present a technique for bidding on multiple Telecom parameters that efficiently trades-off *application relevance* (to increase footfall) with *advertisement-driven monetization*.
- We evaluate our approach using real data from a prominent Telecom operator. The results of our simulation indicate significant improvement in sale of applications for this setting.

The rest of the paper is organized as follows. Section II presents the design of TappStore, including the promotions and bidding platform. Section III is geared towards creating an environment for investigating the performance of TappStore, while Section IV provides initial results based on simulations. Section V discussed related work and finally, we conclude the paper in Section VI.

II. TappStore Design

Design of *TappStore* can be broken down into three main parts. In the first part, we characterize the Telecom data that can be used for promoting applications to consumers in the context of a Telecom Application Store, and give the overall architecture of our TappStore system. In the second part, we give techniques to utilize this data effectively for promoting applications and in the third part, we describe a bidding platform that allows an application provider to place bids on various Telecom information to increase the chances of their applications getting highlighted during search by a user or during cross-sell and up-sell.

A. Overall Architecture

A Telecom operator has various information pieces that are not only *rich*, but also *exclusive* in nature, i.e. this information is not available with other Application Stores such as those hosted by Apple and Android. The key is to use it in conjunction with other information such as application meta-data to promote the most relevant applications to users.

1) Location & Presence Zones: Location information of subscribers is available with a Telecom operator through techniques like Cell tower triangulation, GPS, etc. and has been classically used to offer a number of location based applications and advertisements [1], [2]. In this paper, we propose to represent location information at several different granularities -

- *One or more countries*
- *One or more cities*
- *One or more areas within a city*
  - *<Center point (longitude and latitude), Neighborhood definition (radius in kilometers)>*

Further, TappStore extends the notion of location to include *Presence Zones*. Intuitively, a presence zone is used to detect 'short-term' or 'long-term' stay of a user in a particular location and is formally represented as follows -

<Current Location (as defined above), Duration of stay in this location>

Through presence zone information obtained over a period of time, our intention is to identify base location of users (where they spend maximum time) and also their travel patterns, frequently visited places, etc. It is important to note that while application download and usage may be linked to particular location-time combinations (e.g. user at home in the evening), some applications such as travel based ones may have higher downloads by users who are traveling. Similarly, users might download new applications in their spare time while being away from home, office, etc. In these situations, presence zones become useful.

Location and presence information from Telecom operator is utilized in TappStore from the perspective of the consumer as well as the application. More specifically, for the consumer, this information is used to compute the locations where a user mostly downloads the applications, influence of travel, etc. On the other hand, for the applications, this data is helpful in analyzing the locations where a particular application is mostly downloaded, and that whether downloads occur while users are away from their base location, etc.

2) Telecom Social Networking: Many consumers are influenced by their friends. In essence, a consumer is more likely to buy an application that is already bought by her friends for two reasons (i) the consumer is likely to share interests with a friend, and (ii) the friend might recommend the application (s)he is using. A Telecom operator has massively rich source of information to help find friends of a user – Telecom data of calls and SMSes; this can be mined for revealing calling and messaging patterns.

Similar to [3], we propose that the first step to use this data is to represent it as a *Social Network graph*. This is a directional graph with nodes representing users identified by their mobile number. Whenever there is a call between any two users, an edge is created between the corresponding nodes. The weight of an edge will represent the strength of relationship between the corresponding users calculated using information like total call duration and the frequency of calling and SMSes. Based on analysis of this graph, we form a *friend list* for each user. A friend list is represented as $<(friend1, w_1), (friend2, w_2), ..(friendn, w_n)>$, where w_1, w_2, .. w_n is the weight of the corresponding edge.

Note that for a consumer, such social information is used to compute the influence of friends on her download patterns. On the other hand, for an application, this information is used to ascertain whether downloads occur by a group of friends together, indicating the inherent social nature of the application.

3) Telecom Profile: We consider two other information pieces that, like Location, Presence and Social Networking, are available solely with a Telecom operator.

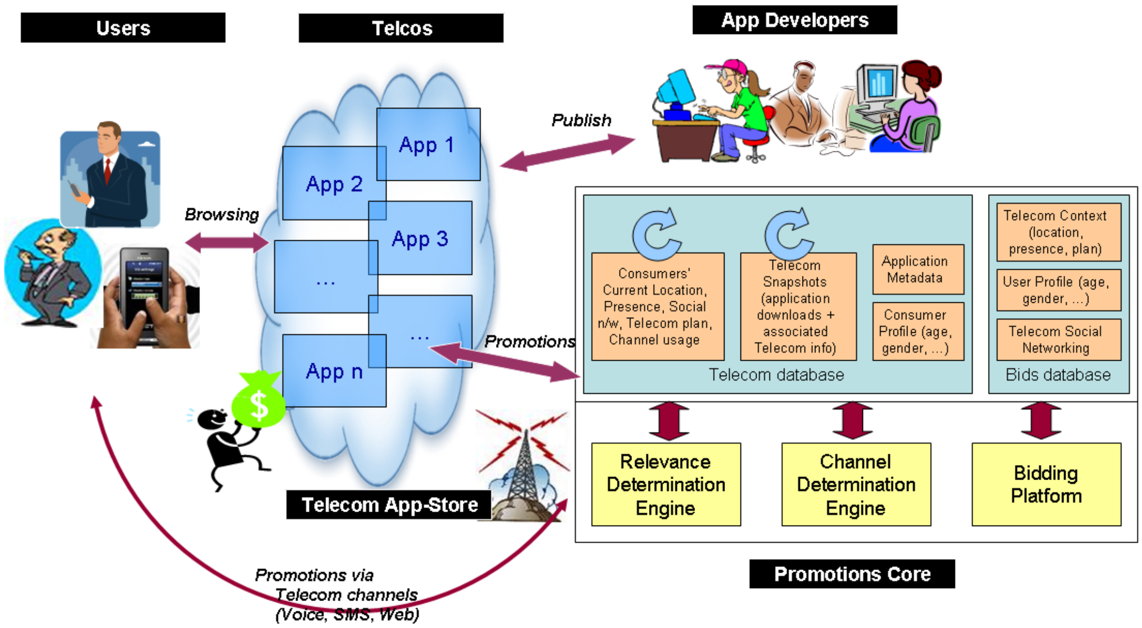

Fig. 1. TappStore Architecture

- Telecom Plan - Application usage may also be correlated with Telecom plan of a consumer. For instance, users with high data plan may use Web-based applications more. Similarly, users with intensive call and SMS plans may use social applications more.
- Telecom Channel Usage - Relative usage of SMS, Call or Web channel varies significantly for users. In TappStore, we propose to incorporate this channel usage information and help advertisers select the best channel from the following - SMS ad, promotional call, a call-back tune or Web advertisement, thereby providing more visibility to promotions.

Telecom information such as current location, presence and social networking of a user are highly dynamic in nature. Therefore, to associate an application download with user Telecom data (for analysis later), TappStore takes a *Telecom Snapshot* at the time of download:

<user_id, application_id, timestamp> <user_location, user_presence, user_friend_list, user_Telecom_profile>

Figure 1 presents the TappStore architecture integrating Telecom data and Telecom Snapshots with various entities in a Telecom Application Store through a *Promotions Core*. This core has the following components -

- Telecom Database. It contains meta-data of applications, each consumer's profile, her Telecom information (location, presence, data plan, Telecom social networking etc.) as well as various snapshots taken during application downloads. By it's very nature, this database is dynamic with continuous updates.
- Relevance Determination Engine. Using the Telecom database, it determines which applications to advertise to a specific user and similarly, to which users to promote a given application. Underneath, this component uses the match-making methodology discussed in the next subsection.
- Channel Determination Engine. Using channel usage information, it determines the best channel for promoting an application to a user.
- Bids Database: It contains bids placed by various developers for different applications. The bids can be placed on Telecom context of a user (location, presence, and Telecom plan), User profile information (age, gender, interests, etc.), and Telecom social networking information (calls made and received, SMS sent and received).
- Bidding Platform: It enables a 'combinatorial' bidding process by allowing developers to place bids on multiple parameters (for promoting an application) and then executes an auction algorithm to decide which advertisement (for an application) to show first, as described in Section II-C. It efficiently trades of *application relevance* (to increase footfall) with *advertisement-driven monetization* to increase the overall revenue of the Telecom operator.

B. Promotions Methodology

The first step in promoting applications is to create *relevance matrices* and associated *metadata table* for both applications and consumers, based on the snapshots taken during various application downloads. As shown in Figures 2 and 3, the application and user relevance matrices give the influence of each dimension on the download of an application by various users, and on the behavior of a user with respect to downloading and using various applications, respectively. The corresponding metadata tables contain additional information that gives further insights on how each dimension influences

Applications	Relative Weights			
	Location	Presence	Social Meter	Purchase History
Application 1	w_{1l}	w_{1p}	w_{1s}	w_{1h}
Application 2	w_{2l}	w_{2p}	w_{2s}	w_{2h}
…	…	…	…	…
Application *n*	w_{nl}	w_{np}	w_{ns}	w_{nh}

Application Relevance Matrix

	Location	Presence	Social Meter	Purchase History
Application 1	$\{<loc_{11}, w_1>, <loc_{12}, w_2>, \ldots\}$	$\{<p_{11}, w_1>, <p_{12}, w_2>, \ldots\}$	$\{<grp_{11}, w_1>, <grp_{12}, w_2>, \ldots\}$	$\{<plan_{11}, w_1>, <plan_{12}, w_2>, \ldots\}$
Application 2	$\{<loc_{21}, w_1>, <loc_{22}, w_2>, \ldots\}$	$\{<p_{21}, w_1>, <p_{22}, w_2>, \ldots\}$	$\{<grp_{21}, w_1>, <grp_{22}, w_2>, \ldots\}$	$\{<plan_{21}, w_1>, <plan_{22}, w_2>, \ldots\}$
…	…	…	…	…
Application *n*	$\{<loc_{n1}, w_1>, <loc_{n2}, w_2>, \ldots\}$	$\{<p_{n1}, w_1>, <p_{n2}, w_2>, \ldots\}$	$\{<grp_{n1}, w_1>, <grp_{n2}, w_2>, \ldots\}$	$\{<plan_{n1}, w_1>, <plan_{n2}, w_2>, \ldots\}$

Application Metadata Table

Fig. 2. Application relevance matrix and the corresponding application metadata table

Users	Relative Weights			
	Location	Presence	Social Meter	Purchase History
User 1	w_{1l}	w_{1p}	w_{1s}	w_{1h}
User 2	w_{2l}	w_{2p}	w_{2s}	w_{2h}
…	…	…	…	…
User *n*	w_{nl}	w_{np}	w_{ns}	w_{nh}

User Relevance Matrix

	Location	Presence	Social Meter	Purchase History
User 1	$\{<loc_{11}, w_1>, <loc_{12}, w_2>, \ldots\}$	$\{<p_{11}, w_1>, <p_{12}, w_2>, \ldots\}$	$\{<frnd_{11}, w_1>, <frnd_{12}, w_2>, \ldots\}$	$\{<profile_{11}, w_1>, <profile_{12}, w_2>, \ldots\}$
User 2	$\{<loc_{21}, w_1>, <loc_{22}, w_2>, \ldots\}$	$\{<p_{21}, w_1>, <p_{22}, w_2>, \ldots\}$	$\{<frnd_{21}, w_1>, <frnd_{22}, w_2>, \ldots\}$	$\{<profile_{21}, w_1>, <profile_{22}, w_2>, \ldots\}$
…	…	…	…	…
User *n*	$\{<loc_{n1}, w_1>, <loc_{n2}, w_2>, \ldots\}$	$\{<p_{n1}, w_1>, <p_{n2}, w_2>, \ldots\}$	$\{<frnd_{n1}, w_1>, <frnd_{n2}, w_2>, \ldots\}$	$\{<profile_{n1}, w_1>, <profile_{n2}, w_2>, \ldots\}$

User Metadata Table

Fig. 3. User relevance matrix and the corresponding user metadata table

the decision.

1) Application Relevance Matrix: For each application, the relevance matrix (refer Fig. 2) contains relative weights for various dimensions that influence purchase behavior of consumers. In particular, we first determine the weight of each dimension and then compute relative weights across various dimensions.

- Location - If an application has been downloaded and used many times from some location(s), then a high weight is assigned to this dimension in the relevance matrix and the top download locations are stored in the application metadata table. The weight of each location (based on the number of times an application is downloaded from that location compared to total downloads of that application) is also stored with each entry as shown in Fig. 2.
- Presence - If an application has been downloaded and used by many users from a location where they have a short term (or long term) presence, then a high weight is assigned to this dimension and the required presence information is associated with the application in the application metadata table, similar to location information.
- Social Meter - If an application is found to be used in social groups, then a high weight is assigned to this dimension and the top groups (whose maximum number of users have downloaded the application) are stored in the application metadata table. The weight of each group, based on how many users within the group have downloaded the application, is also stored.
- Purchase History - If an application has been downloaded and used by many users with a particular Telecom data or call plan, then a high weight is assigned to this dimension and the top plans and their weights are stored in the metadata table.

2) User Relevance Matrix: This matrix (refer Fig. 3) consists of relative weights for various dimensions that influence purchase behavior based on user's dynamic profile and past purchases. The corresponding user metadata table contains further information about each dimension and how it influences the user's purchase decision.

- Location - If the user downloads many applications from particular location(s), a high weight is given to this dimension in the relevance matrix and the top locations are associated with the user in the user metadata table. The weight of each location (based on number of applications downloaded from this location compared to total applications downloaded by this user) is also stored with each entry.
- Presence - If short or long term presence of user in various zones has influenced her purchase decisions in the past, high weight is assigned to this dimension. Corresponding entries are made in the user metadata table, similar to location information.
- Social Meter - A weight proportionate to the influence of friends on user's download pattern (determined by whether the user has downloaded the same applications as her friends) is stored in the relevance matrix. User metadata table in this case contains individual friends and their corresponding influence.
- Purchase History - If the user has in past downloaded many applications with particular profiles (specific categories, etc.), then a high weight is assigned to this

dimension in the relevance matrix. The top application profiles and their weights are stored in the metadata table.

Now, there are two different use cases for advertising applications – 1) *promotion* i.e., selecting the most relevant application to advertise to a given user, and 2) *campaign* i.e., finding a set of 'k' best users to whom a given application should be promoted. Below, we describe the matchmaking algorithms for these two cases:

Promotions - This pertains to deciding which applications to advertise to a particular consumer. TappStore utilizes the following algorithm:

Step 1. *For a user 'c', determine relative weight w_{ci} for each dimension 'i' in the user relevance matrix, where $i \in \{location, presence, social\ meter, purchase\ history\}$.*
(Calculated by analyzing influence of this dimension on application purchase decision, as described before)

Step 2. *Determine overall propensity of the consumer to purchase applications w_{cp}*
(Calculated as a function of applications browsed and downloaded, normalized to be a value between 0 and 1. We describe this notion further in Section III)

Step 3. *Calculate relevance score R_{ca} of each application 'a' for this consumer*
$R_{ca} = \Sigma(w_{ci} * s_{ia})$*, where s_{ia} is a normalized value which indicates satisfaction of dimension 'i' for this application*
(For each dimension, different functions are used to determine the satisfaction. For example, satisfaction of the 'social' dimension is calculated as a function of how many friends have downloaded this particular application and the influence of those friends on the consumer (determined by the weight of the corresponding edge in the Social Network graph). Similarly, purchase history of the application and the consumer is used to give high scores to applications similar to ones the consumer has purchased before or that have been purchased by users of similar Telecom plans)

Step 4. *If $(R_{ca} * w_{cp}) >$ Threshold 'T_c', promote this application*
(Intuitively, this considers both the propensity of consumer to purchase applications as well as relevance of the application in question)

Campaigns - This corresponds to deciding which consumers to promote a given application. TappStore uses the following algorithm:

Step 1. *For an application 'a', determine relative weight w_{aj} for each dimension 'j' in the application relevance matrix, where $j \in \{location, presence, social\ meter, purchase\ history\}$.*
(Calculated by analyzing influence of this dimension on application purchase decision, as described before)

Step 2. *Determine overall popularity of this application w_{ap}*
(Calculated as a function of total downloads, normalized to be a value between 0 and 1. We describe this notion further in Section III)

Step 3. *Calculate relevance score R_{ac} of each consumer 'c' for this application*
$R_{ac} = \Sigma(w_{aj} * s_{jc})$*, where s_{jc} is a normalized value indicating satisfaction of dimension 'j' for the consumer*
(For example, satisfaction of the 'presence' dimension is determined by evaluating the current presence zone of the consumer and that whether the application is suitable for that presence zone. Similarly, satisfaction of 'social' dimension is based on the number of friends of the consumer who have downloaded this application and the influence of those friends on the consumer.)

Step 4. *If $(R_{ac} * w_{ap}) >$ Threshold 'T_a', promote to this consumer*
(Intuitively, this considers both the application popularity as well as relevance of this application to a consumer)

C. Bidding Platform

Algorithms used by e-commerce sites, such as e-Bay and Amazon, are designed for promoting products, do cross-selling and up-selling based on consumer's current search, popularity of different products, and/or profile of the consumer. On the other hand, search engines, such as Google, allow advertisers to place bids on search keywords. Here, advertisements are shown to a user based on keyword matching. Even though there is no bidding on geographical location, advertisers can specify location where advertisements would be deemed appropriate.

With thousands of applications being possibly published monthly in a mobile application store, it becomes imperative from the developers' perspective that their applications be highlighted during search by consumers as well as for cross-sell/up-sell. Like with promotions discussed in the previous subsection, we argue that a whole new eco-system is possible between Telecom operators, application developers and consumers where an Application Store is monetized through 'bidding' on different Telecom information pieces by developers.

TappStore offers a bidding platform to developers. In essence, this platform enables developers to define relevance criteria for each information piece, and thereafter utilizes a 'combinatorial' mechanism to incorporate bidding based on individual information pieces, efficiently doing a trade-off between monetary bids, profits from sale of applications, as well as relevance, to advertise applications to consumers.

The overall methodology for bidding in TappStore is based on enabling the application developer to select a combination of *filter criteria* and *bidding parameters* from the set of available Telecom parameters. The filter criteria is geared towards establishing basic relevance between consumers and applications, for example, suitability of application to a particular geographical location only. The bidding parameters, on the other hand, serve as a tool to mathematically capture a developer's perception of application vis-a-vis the consumers in a Telecom eco-system. The developer specifies total bid value for promoting her application to consumers, with relative weights ascribed to different Telecom parameters around which the bid is placed.

Filter criteria as well as bids on various Telecom parameters are entered by a developer using the *Bidding Console* of TappStore, as shown in Figure 4. In the particular example captured in this figure, we see that the developer has selected a filter criteria around location and Telecom data plan, while

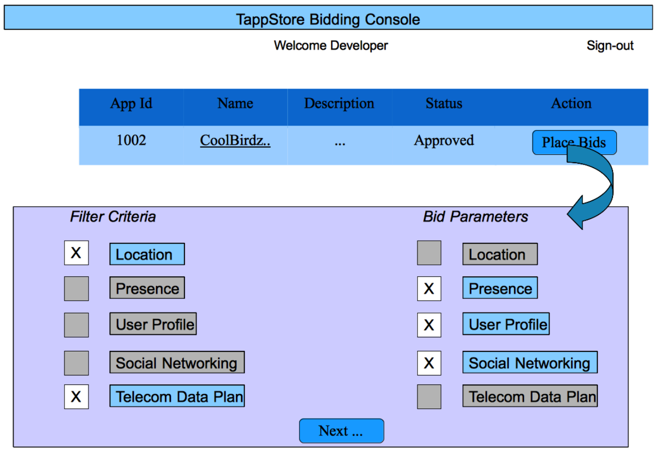

Fig. 4. TappStore Bidding Console

placing bid on the other parameters. Figure 5 presents the flow of the bidding process, involving details of the filter criteria as well as the bid parameters, and specification of the overall bid value and weights of individual parameters. As shown, for location, the advertiser needs to qualify whether it corresponds to a country, or a city, or specific region(s) within a city described using a center point (latitude, longitude) and neighborhood (radius). Similarly, for social networking information, one has to specify the criteria for determining friends, using the Calls and/or SMS information, as explained earlier. Moreover, 'minimum number of friends' of this user who should have downloaded this application is also provided.

Matchmaking Algorithm. TappStore determines the best application to be shown to a consumer, based on a trade-off between bid and relevance:

Step 1. *Gather consumer's real time information (location and presence)*

Step 2. *Get consumer's other information (customer profile, activity history, social networking)*

Step 3. *Use filter criteria to short-list appropriate applications*

Step 4. *Calculate relevance score R_{ca} of each application 'a' for this consumer 'c'*

$R_{ca} = \Sigma(w_{ci} * s_{ia})$, where s_{ia} is a normalized value which indicates satisfaction of dimension 'i' for this application

(Note that this step is similar to the one in 'Promotions')

Step 5. *'Boost' R_{ca} with the normalized (i.e. with respect to other applications) profit p_a that the operator would make with sale of this application to obtain R_{cab}*

$R_{cab} = R_{ca}/(1 - p_a)$

(Intuitively, this incorporates the inherent 'monetary' value of the application 'a' for TappStore)

Step 6. *Compute score S for each application based on the bid value and weights assigned by the developer*

$S = B*(w_1*ps_1 + ... + w_n*ps_n)$, where B is the total bid value of the application, w_i is the weight for parameter i and ps_i is the boolean indicating whether this consumer satisfies the parameter i

Step 7. *Compute an overall ranking of appropriate applications, based on R_{cab} and S for each application*

(Intuitively, this incorporates a trade-off between relevance of application and expected gains to TappStore through the bidding process)

Step 8. *Run an auction-based algorithm [2] to find cost that the top advertisers need to pay*

(Essentially, this compares the multiple bids by various advertisers to determine the best offers and the associated bid cost)

Note that this match-making could be used beyond applications for general advertisement also. In that case, application metadata and download history would not be used, and there would be no boosting in Step 5.

III. TappStore Evaluation

Our end goal is to evaluate TappStore in a real Telecom setting. However, till the point we reach that stage, we make an attempt to asses TappStore using two models based on real-world data. The first model is that of an Application Store with various facets derived from statistics available for Android Marketplace. The second model pertains to consumers (their social structure, purchase behavior, etc.) derived using data for 28 million users from a large Telecom operator in Asia. Our objective here is to compute the increase in expected sale of applications in a simulated Telecom Application Store utilizing the above models.

A. Application Store

In an Application Store, some applications are more popular than others and have higher probability of being downloaded. The applications can also be grouped into different categories. For these, we followed Android applications' popularity and category distribution, and also modeled the download of applications by various users.

1) Popularity Distribution: We take the popularity of any application to be proportional to its number of downloads, relative to other applications in the store. Formally, relative popularity is defined as

$$Rel.\,popularity = \frac{no.\ of\ downloads\ of\ the\ application}{max.\ no.\ of\ downloads\ of\ any\ app\ in\ the\ app-store} \quad (1)$$

We obtained the download distribution of applications in Android Marketplace from *http://www.androlib.com/appstats.aspx*. While simulating a given number of applications in our Application Store, we fitted the download distribution of Android applications to this number and computed the relative popularity values.

2) Category Distribution: Applications can be broadly divided into various categories such as sports, education, social, travel, etc. For a give number of applications in our store, we use the same category distribution as in Android Marketplace (obtained from *http://www.appbrain.com/stats/android-market-app-categories*).

[2]en.wikipedia.org/wiki/Auction_algorithm

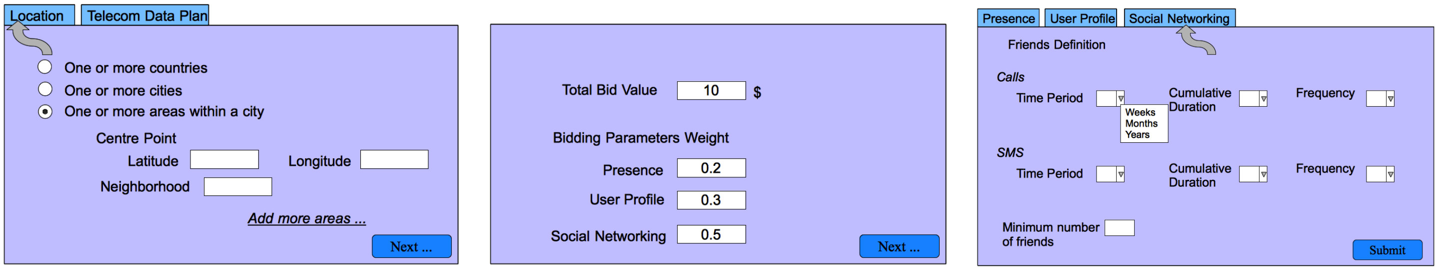

Fig. 5. Flow of the Bidding Process in TappStore (a) Details of Filter Criteria (b) Specification of Bid Value and Weights of individual Parameters (c) Details of Bid Parameters

3) Application Downloads: To model the download of applications by users, we used real data of purchase by Android phone users. On average, an Android user has 40 applications. Also, a user downloads 1 paid application per 8 free applications, implying that she uses 4.4 (i.e. $40/9$) paid applications on an average. Hence, we used a normal distribution with a mean of 4.4 (standard deviation of 2, restricting the values between 0 and 8) to obtain distribution of number of applications purchased by different users. For simulation purposes, this distribution was mapped to a given number of users. Assignment of specific applications from the Application Store to users was done randomly, in proportion to their relative popularity values.

B. Consumers

For consumers, we modeled the users' propensity to purchase and their social structure using real data from a large Telecom operator in Asia (identity concealed due to contractual obligations).

1) Propensity to Purchase: Intuitively, this is defined as an individual's tendency to purchase an application that is advertised to her. Formally, we define propensity to purchase as follows:

$$Propensity\ to\ purchase = \frac{No.\ of\ applications\ downloaded}{Total\ no.\ of\ known\ (or\ advertised)\ apps} \quad (2)$$

While creating this model, we were hampered by lack of directly relevant data from a Telecom Application Store. To work around this limitation, we looked at data on visits by 2 million users to 21 websites on mobile phones, obtained from the aforementioned Telecom operator. We argue that since most of the applications are downloaded online from an Application Store, mobile internet usage indicates activity of a user beyond just calls and SMSes and can be used to roughly estimate propensity to purchase values of Telecom subscribers.

In specific, we treated websites as mobile applications and visit to a website as download of that application. We assumed that the 21 websites that the 2 million users visited in all are popular and known to all in general. Using this, we approximated the propensity to download an application on seeing its advertisement by using the following formula:

$$Propensity\ to\ purchase = \frac{Number\ of\ websites\ visited}{Total\ number\ of\ popular\ websites} \quad (3)$$

We found a percentage-wise distribution of propensity values for 2 million users and fitted it to a given number of consumers during our simulation.

2) Social Structure: We collected month-long data on calls and SMSes of 28 million users from our Telecom operator. From this data, we created the Social Network Graph and formed the friend list for each user as described in Section II. Before forming the friend list, we removed nodes which had a high number of incoming edges with very few outgoing edges (most probably delivery centers, etc.) as well as nodes which had a high number of outgoing edges with very few incoming edges (most probably Telemarketers). After forming the friend lists for all the 28 million users, like with other distributions, we mapped this distribution to a given number of users required for a simulation.

C. Approach for Simulation

During a simulation, we advertise applications to users. Depending on a user's propensity to purchase, application's popularity and 'appropriateness' of the application for the user, we estimate the user's probability to buy the application. By summing this probability over all users, we obtain the expected sales. We consider two cases – one, where Telecom information is used to advertise relevant applications to a user, and two, where general advertisement is done without using any Telecom information. We compared the expected sales for the two cases and computed the increase in sale due to use of Telecom information. In both cases, we use the following formula for probability of purchase

$$P(U_i, A_j) = \kappa * Pr(U_i) * R\_Pop(A_j) * A - factor \quad (4)$$

where
$P(U_i, A_j)$**: Probability of purchase by user i on being advertised application j.**
κ **: Constant of proportionality**
$Pr(U_i)$ **: Propensity to purchase of user i**
$R\_Pop(A_j)$ **: Relative popularity of application j**
$A - factor$ **: Appropriateness factor**

$$Expected\_Sales = \Sigma_i P(U_i, A_j) \quad (5)$$

$$A - factor = \{^{> \ 1 \ (varied\ from\ 2\ to\ 5)\ for\ Telecom\ hit}_{= \ 1 \ otherwise} \quad (6)$$

Intuitively, *Telecom hit* in equation 6 increases the probability of purchase by increasing appropriateness of applications to users and occurs when one or more of the following conditions are satisfied:

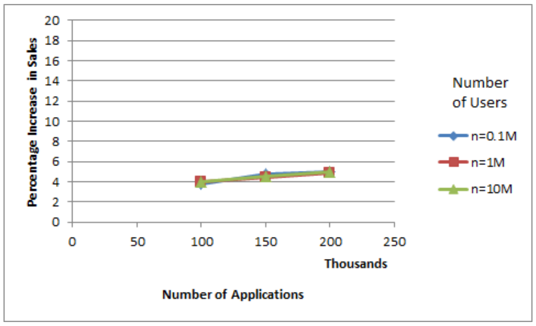

Fig. 6. Increase in sales of applications using location information

- A travel-based application is advertised to a user who is currently traveling or is in a new place
- A user is advertised an application also being used by her friends
- A socially active person is advertised a social application

Note that this is in line with the campaign and promotion algorithms presented in Section II. For Telecom case, in the current implementation, we advertise applications using location/presence and social data. We do not consider other Telecom information related to Telecom channel and plan, primarily due to lack of availability of the corresponding data from our Telecom operator.

Hill et al. [4] reported that *Network neighbors* - those customers who communicated with a prior customer - adopt a service at a rate 3-5 times greater than baseline groups selected by the best practices of a firm's marketing team. Hence, we varied the appropriateness factor from 2 to 5 in steps of 1 in our experiments. κ in equation 4 gets canceled in computing percentage increase of sales in the Telecom case over the random case. We have varied the number of applications in our experiments around the actual number of applications in Android Marketplace, from 100,000 to 200,000. For users, we varied the number from 0.1M to 10M.

IV. Simulation Results

We conducted various simulations to determine the increase in sale of applications for the two uses cases of promotions and campaigns; results are described below. We intend to perform similar experiments for the bidding platform in the near future.

A. Promotions

In this experiment, each user is advertised an application. For Telecom case, selection of application for promotion depends on application's popularity as well as Telecom data such as social structure and location. For the other case, we randomly select an application solely based on its popularity.

1) Using Location Information: Figure 6 presents the positive influence of location dimension on promotions. Herein for the Telecom case, we advertised travel-based applications to people who are currently traveling outside their base locations and used an appropriateness factor of 2. Note that increase in sales is independent of the number of users. However, as number of applications increases, increase in sales goes up because the number of high popularity travel applications goes up. So for the Telecom case, more popular travel applications are advertised to travelers.

2) Using Social Networking Information: Figure 7 shows the effect of social dimension on promotions. In particular, for Telecom case, we advertised that application to a user which is also bought by her friends and further evaluated two approaches to select such an application. Our first approach is to select randomly from the set of applications bought by friends and keep appropriateness factor constant at 2 (Fig. 7(a)). In the second approach, we pick the application that maximum number of friends use. Here, we vary the appropriateness factor from 2 to 5, in proportion to number of friends using the application (Fig. 7(b)).

Since number of users is very large as compared to number of friends of any user, the friends list size remains same even with more users. Hence, the percentage increase in sales is independent of the number of users. In approach 1, since number of high popularity applications increases with increase in total number of applications, there is an increase in sales as number of applications goes up. However, in approach 2, an opposite trend is observed. This is because appropriateness factor varies with number of friends sharing an application. As number of applications increases, the number of friends sharing an application goes down, decreasing the appropriateness factor in the Telecom case. Observe that the overall increase in sales is significantly higher in approach 2 compared to the first one, as promoting the application used by most friends of a user results in higher probability of download.

B. Marketing Campaign for a New Application

In this experiment, we advertise a new application to k best users. We compare the sales for approaches using and those not using Telecom data. For the case without using Telecom data, we select k users with maximum propensity to consume the advertised application. Let us look at the results for two sub-cases where the new application is either travel based or a social application.

1) Travel-based application: For marketing a new travel-based application, we select k customers from set of travelers with highest propensity to purchase utilizing the Telecom information of short-term presence. Figure 8(a) shows the expected improvement in sales using Telecom data with an appropriateness factor of 2. From the figure, we also see that the improvement in sales increases slightly as number of users grow, as there are more travelers in the set of users with high propensity to purchase.

2) Social application: For marketing a new social application, we select k customers from the set of socially active people with highest propensity to purchase, based on the Telecom social networking data. A person is socially active if she calls or messages more than 10 people in a month. Figure 8(b) presents the corresponding improvement in sales using Telecom data. Similar to the travel case, increment in sales grows slightly with number of users as there are more

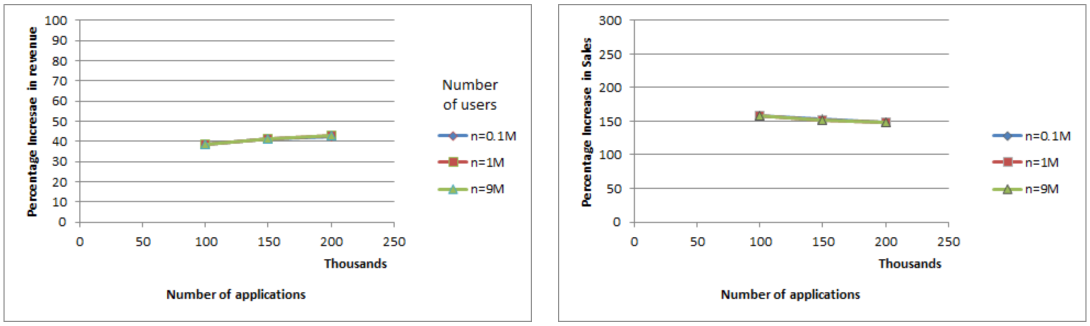

Fig. 7. Increase in application sales using social networking information: (a) Approach 1 (b) Approach 2

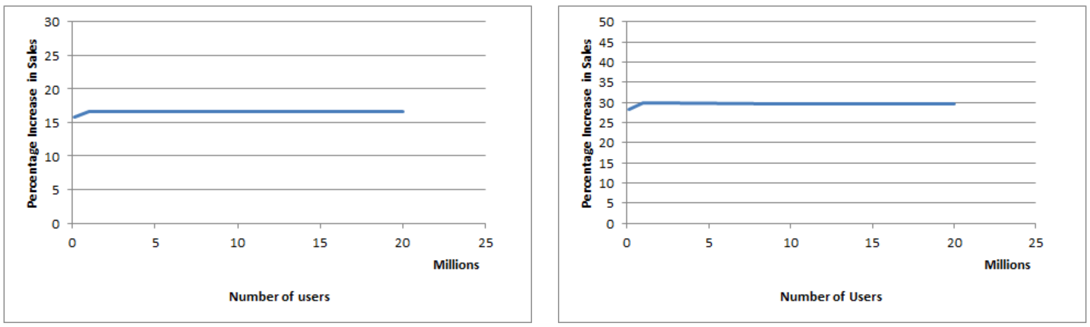

Fig. 8. Best 'k' campaign for a new (a) travel application (b) social application

number of socially active people in high propensity to purchase region.

V. Discussion and Related Work

E-commerce websites normally use customers' history, age, gender, product popularity and sellers' ratings for promoting the most relevant products. Wu et al. [5] predicted the conversion probability for items on C2C E-commerce sites. They used product specific features like *category hotness*, *item hotness*, price, number of competitors, conversion rate of similar items, etc. and seller specific features, such as seller feedback score, to rank search results. Parikh et al. [6] created a user-tunable model for ranking marketplace search results based on linear combination of relevance, diversity, trust and value. Speretta et al. [7] built individual user profiles based on past search activities to provide personalized search results, obtaining a 34% improvement in the rank-order. Our approach for mobile application marketing in TappStore involves the widely-used Telecom setup to gain deeper insights into dynamically changing user profile based on current location, social circle and other Telecom data. However, the above cited works can complement TappStore and help incorporate other parameters such as age, gender and past behavior.

The role of social networking in e-commerce has been studied in great detail. Hill et al. [4] found that statistical models for promotions, built with geographic, demographic and prior purchase data are significantly improved by including network information. Guo et al. [8] by studying triads and the directed closure process, quantified the presence of information passing. They determined that social graph is a far better determinant of consumer behavior than meta-data features such as seller reputation or product price. Similarly, [9] shows how viral marketing can be propagated in a social network by targeting some customers with high network value. These studies not only underscore the importance of social networking information, but can also be utilized in TappStore to complement our current approach of advertising a mobile application to friends of a customer.

The number of applications in app-stores by device vendors like Android and iPhone is increasing day by day. Android Marketplace has over 200,000 applications. Android deploys user rating to rank applications. AppAware [10] is a mobile platform that provides a way to exploit installations and removals of Android applications as implicit ratings to define an application community perceived quality. One key difference between Application Stores by device vendors and those by Telecom operators is that, unlike device vendors who have to provide applications only for their own platform, Telecom operators serve customers with multiple device platforms. The Wholesale Applications Community (WAC) [11] is an organization that proposes a unified and open platform to allow mobile software developers to write applications for a variety of devices, operating systems and networks. Our work can help efforts like WAC compete effectively with device vendors hosted Application Stores by helping promote applications efficiently.

Advertising based on keywords used in search query gives rise to "sponsored links" that appear alongside search results, and has grown into a multibillion-dollar market. For example, one of Google's sources of revenue is its AdWords[3] where advertisers bid on keywords. A number of research groups have studied the problem of bidding and auctioning on keywords. Chen et al. [12] look into the "share structure" problem - that is, of the total available resources for each keyword, how large a share should be set aside for the highest bidder, for the second-highest bidder, and so on. In [13], the authors propose the use of hybrid auctions where an advertiser can make a per-impression as well as a per-click bid, and the auctioneer then chooses one of the two as the pricing mechanism. [14] examines a model in which position auctions influence consumer search. We wish to highlight that situation in TappStore is more complex as it allows bidding on multiple dimensions, not only keywords. However, going forward, we intend to explore the applicability of works such as those cited above to our bidding platform.

The issue of consumer privacy is inherent to the domain of targeted promotions, and has been analyzed for Web [15], [16], as well as M-Commerce [17] extending to mobile applications [18], [19]. In [20], the authors conduct a large scale study on the risk of re-identification attacks with published location data obtained through call data records. Since TappStore also makes use of consumer data, privacy issues apply here too. In the short term, we plan to provide options to consumers whereby they can specify whether they are comfortable or not with the notion of TappStore utilizing their information for promotions. A comprehensive treatment to this topic, however, is the scope of future work.

VI. Conclusion

Telecom operators are currently side players in the mobile applications space and are treated as mere dumb pipes by users for accessing these applications. However, they have some core strengths which can make them significant players in this space. In this paper, we proposed an architecture for a Telecom Application Store (TappStore) that utilizes data available only in a Telecom setting, such as user location, presence, social networking derived from calling patterns, and Telecom profile to effectively promote relevant applications to users. TappStore also allows application providers to bid for various Telecom information pieces to better promote their applications. Initial evaluation based on simulations using real data from a large Telecom operator demonstrate that our approach can lead to increased sale of applications in a Telecom Application Store.

In the future, we would like to do a real pilot with a Telecom operator for application promotion to gain more insights and further improve our methodology. Moreover, we would like to extend TappStore in several aspects, such as enabling privacy policies for consumers.

Acknowledgment

The authors would like to thank the 'SNAZZY' team at IBM Research - India, for helping them with the data during the simulation process.

[3] https://adwords.google.co.in/

References

[1] D. Chakraborty, K. Dasgupta, S. Mittal, A. Misra, A. Gupta, E. Newmark, and C. L. Oberle, "BusinessFinder: Harnessing Presence to enable Live Yellow Pages for Small, Medium and Micro Mobile Businesses," *IEEE Communications Magazine*, vol. 45, no. 1, January 2007.

[2] "VZ Navigator," http://www.vznavigator.com/.

[3] A. A. Nanavati, R. Singh, D. Chakraborty, K. Dasgupta, S. Mukherjea, G. Das, S. Gurumurthy, and A. Joshi, "Analyzing the Structure and Evolution of Massive Telecom Graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 703–718, 2008.

[4] S. Hill, F. Provost, and C. Volinsky, "Network-Based Marketing: Identifying Likely Adopters via Consumer Networks," *Statistical Science*, vol. 21, pp. 256–276, 2006.

[5] X. Wu and A. Bolivar, "Predicting the Conversion Probability for Items on C2C Ecommerce Sites," in *Proceeding of the 18th ACM Conference on Information and Knowledge Management*, Hong Kong, China, 2009.

[6] N. Parikh and N. Sundaresan, "A User-Tunable Approach to Marketplace Search," in *Proceedings of the 20th International Conference Companion on World Wide Web*, ser. WWW '11, Hyderabad, India, 2011.

[7] M. Speretta and S. Gauch, "Personalized Search Based on User Search Histories," in *Proceedings of the International Conference on Web Intelligence*, 2005.

[8] S. Guo, M. Wang, and J. Leskovec, "The Role of Social Networks in Online Shopping: Information Passing, Price of Trust, and Consumer Choice," in *Proceedings of the 12th ACM Conference on Electronic Commerce*, San Jose, California, USA, 2011.

[9] M. Richardson and P. Domingos, "Mining Knowledge-Sharing Sites for Viral Marketing," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, 2002.

[10] A. Girardello and F. Michahelles, "Bootstrapping Your Mobile Application on a Social Market," in *Workshop "Research in the Large" at Ubicomp*, Copenhagen, Denmark, September 2010.

[11] "Wholesale Applications Community," http://en.wikipedia.org/wiki/Wholesale_Applications_Community.

[12] J. Chen, D. Liu, and A. B. Whinston, "Auctioning Keywords in Online Search," *Journal of Marketing*, vol. 73, no. 4, pp. 125–141, 2009.

[13] A. Goel and K. Munagala, "Hybrid Keyword Search Auctions," in *Proceedings of the 18th international conference on World Wide Web*, New York, NY, USA, 2009, pp. 221–230.

[14] S. Athey and G. Ellison, "Position Auctions with Consumer Search," National Bureau of Economic Research, Working Paper 15253, August 2009.

[15] H. Wang, M. K. O. Lee, and C. Wang, "Consumer privacy concerns about Internet marketing," *Commun. ACM*, vol. 41, pp. 63–70, March 1998.

[16] D. S. Evans, "The Online Advertising Industry: Economics, Evolution, and Privacy," *Journal of Economic Perspectives*, vol. 23, no. 3, pp. 37–60, 2009.

[17] E. Gratton, "M-commerce : The Notion of Consumer Consent in Receiving Location-Based Advertising," *Canadian Journal of Law and Technology*, 2002.

[18] D. Wetherall, D. Choffnes, B. Greenstein, S. Han, P. Hornyack, J. Jung, S. Schechter, and X. Wang, "Privacy Revelations for Web and Mobile Apps," in *Proceedings of the 13th USENIX conference on Hot topics in operating systems*, 2011, pp. 21–21.

[19] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones," in *Proceedings of the 9th USENIX conference on Operating systems design and implementation*, 2010, pp. 1–6.

[20] H. Zang and J. C. Bolot, "Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study," in *Proc. of ACM Mobicom*, Sep. 2011.