Privacy Preserving in Personalized Mobile Marketing

Yuqing Sun and Guangjun Ji

School of Computer Science and Technology, Shandong University
sun_yuqing@sdu.edu.cn, jgj_ak@163.com

Abstract. With the popularity of smart portable devices and advances in wireless technologies, mobile marketing increases quickly. Among various methods, short message is regarded as the most efficient mode. While mobile advertising enhances communication with consumers, the messages without a required permission cause privacy violations. So, how to simultaneously supporting personalization and privacy preserving in mobile marketing is a challenging problem. In this paper, we investigate this problem and propose a Privacy Preserving Model for Personalized Mobile Marketing (P^2PMM), that can protect both users' preferences and location privacy while supporting personalization in mobile marketing. We propose an efficient coding method to manage hierarchical categories and users' preferences. We also investigate how such a model can be applied into practice and implement a prototype.

1 Introduction

Mobile Marketing is a set of practices that enables organizations to communicate and engage with their audience in an interactive and relevant manner through any mobile device or network [1]. With advances in wireless technologies and the popularity of smart portable devices, mobile marketing increases quickly. Among various methods, short message is regarded as the most efficient mode [2] when businesses started to collect mobile phone numbers and send off wanted (or unwanted) content to users. Another trend on mobile marketing is the type of location-based services (LBS) that can offer messages to users based on their current location. The cell phone service providers get the location from a GPS (Global Positioning System) chip built into a phone, or using radiolocation and trilateration based on the signal-strength of the closest cell-phone towers.

While mobile advertising enhances communication with consumers, the messages without a required permission from the consumer cause privacy violations. Actually no matter how well advertising messages are designed, if consumers do not have confidence that their privacy will be protected, this will hinder their widespread deployment[3]. A number of important new concerns emerged mainly stem from the fact that mobile devices are intimately personal and are always with the user, and four major concerns can be identified: mobile spam, personal identification, location information and wireless security [4]. Experts cited fear of spam as the strongest negative influence on consumer attitudes towards SMS advertising [5]. So advertisers should have permission and convince consumers to "optin" before sending advertisements. For example, a simple registration ensures sending relevant messages to an interested audience[6].

However, user choices for interested messages may cause other privacy concerns since tailoring messages requires gathering users' data, such as the preferences, position change or even behavior. Membership systems may reinforce fears of unwanted messages and misuse of personal data. Although there are much research work on either privacy preserving in location based services or personalized mobile marketing [7–9], to the best of our knowledge, there is no literature comprehensively targeting the two competitive objectives of personalization and privacy. In this paper, we investigate the problem on how to protect users' privacy while supporting personalization. In the proposed Privacy Preserving Model for Personalized Mobile Marketing (P^2PMM), users can customize their preferences for messages without any privacy leakage to information provider. The trusted third party collects the marketing messages and make classification according to predefined categories. Users thus have their options on location, time and categories etc. We investigate how such a model can be realized in the GPS and cellular network systems. The prototype system is designed and implemented.

The remainder of the paper is organized as follows. Section 2 outlines related work. Section 3 presents the main components of the proposed model. Section 4 investigates the problem of efficient information organization and query processing in the model. Section 5 discuss the system architecture and describes the details on implementation of the prototype. Section 6 concludes the paper and outlines future research directions.

2 Related Work

Our work is related with the privacy preserving in location based services. Location information is critical for providing customized services, on the other hand, if misused, can lead to privacy breaches[10]. To address such problems, different techniques have been proposed that are based on two main approaches: *location cloaking*, under which a suitable large region is returned to the service provider instead of the precise user location [11]; *location k-anonymization*, under which the location of an individual is returned to the service provider only if it is indistinguishable with respect to the location of other k-1 individuals [7, 8]. However these techniques do not support personalized privacy preferences. Damiani et al. develop a Privacy-preserving Obfuscation Environment system (PROBE)[9], in which spatial entities are divided into two categories: sensitive entities and unreachable entities. PROBE allow individuals to specify the types of entity and risk thresholds as their personal profiles. When requesting LBS, PROBE is able to generate a generalized location so that the probability that an attacker is able to determine the actual individual location is below that threshold. Cheng et al. [12] suggest a framework to provide high quality and privacy-preserving services. However, these methods are not suitable for privacy preserving in mobile marketing since they do not consider customization requirement of message contents.

Another related work is on personalized mobile marketing. Short messages are regarded as the most efficient mode. However, unsolicited messages, commonly known as spam [5], stifle user acceptance. To support personalization, messages should be appropriately tailored before sending to consumers[13, 2]. Advertisers should have permission and convince consumers to "optin" before sending advertisements, such as user registration for interested messages. Solutions have been deployed to personalize text messages based on the consumer's local time, location, and preferences [14], e.g. directions to the nearest vegetarian restaurant open at the time of request. However, personalization in mobile marketing means collecting and storing information about a particular person, such as monitoring of user behavior. This causes privacy concerns. To the best of our knowledge, there is no work that well solve above two challenging objectives of privacy and personalization.

3 The Privacy Preserving Model for Personalized Mobile Marketing

In this section we introduce the Privacy Preserving Model for Personalized Mobile Marketing (P^2PMM for short), depicted as Figure 1. There are four entities in this model. The functionalities of each party are narrated as follows.

- Mobile network operator (MNO): is a telephone company that provides communication services for mobile phone subscribers.
- Intermediary services provider(ISP): is a trusted third party for users that is independent to Merchants. It is responsible for providing the platform for merchants to manage their advertisements, as well as for users to subscribe their interested messages and maintain individual private data. Sometime, it can be integrated with MNO if required.
- Users: are the cell phone subscribers. After registering on the Intermediary services provider, they are allowed to option their preferred messages from ISP based on their location, time or interested topics.
- Merchants: represent the organizations who want to advertise their business messages. After registering on ISP, they are allowed to publish their advertisements to interested users.

There are three distinct characteristics from other mobile marketing models. Firstly, it is active marketing. The P^2PMM model is in a "PULL" schema rather than a traditional "PUSH" way such that all the messages sent to users are what they want. Second is that the users' preferences privacy are preserved. The sensitive information of each individual, such as the profile, the preferences in each query and current location, are stored in the trusted Intermediary services provider, which avoids the case that every merchant has a copy of user profile. Thirdly, users' location privacy are also preserved. The ISP can only acquire the approximate location square without any awareness of the exact real-time position of a user.

3.1 Basic Terminologies

In this section, we would present the formal definition of basic notions in our model. Let U and M respectively denote the set of users and the set of merchants.

Definition 1 (Position). *A geographic position denotes a point in a map and is in form of $Loc = [lngt, lat]$, where $lngt$ and lat are the longitude and latitude value of this point.*

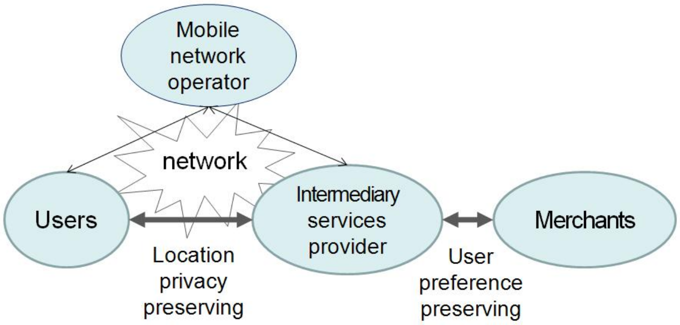

Mobile network operator
network
Users
Intermediary services provider
Merchants
Location privacy preserving
User preference preserving

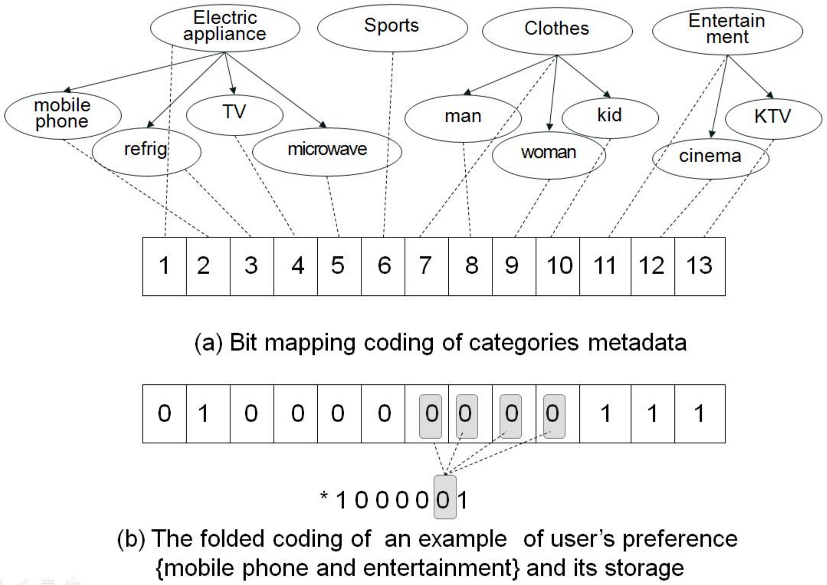
Fig. 1. The Privacy Preserving Model for Personalized Mobile Marketing

Let POS denote the class of positions. We assume that every merchant is associated with an exact *position*. When a merchant registers on the ISP server, its position can be acquired by some position technologies like $GoogleMap$. Similarly, every mobile phone user has an exact *position* at any time, which can be acquired by a GPS or wireless positioning technologies. We introduce two predicates here $LocU(u \in U) : POS$ and $LocM(m \in M) : POS$ to calculate the location of a user and a merchant, respectively.

Definition 2 (Message). *Let $ISSUE$ denote the set of all issues considered in the model. A message msg is specified as a tuple msg=$< ID, TXT, Issues, TW, Loc >$, where ID is the unique identifier of message msg, TXT represents the message content, $Issues \subseteq ISSUE$ is a subset of $ISSUE$ denoting the issues correlated with the message, TW is in form of $[t_1, t_2]$ representing the time window when message msg is effective, and $Loc \in POS$ is in form of $[lngt, lat]$ denoting the merchant's position who launches this message.*

For example, the department store $Macy^*s$ in West Lafayette, IN wants to make a advertisement for sales promotion. It launches the following message to the Intermediary services provider: `msg=<201006081123,``There is a 10% discount on mobile phones for MACY WIP in West Lafayette", {discount, mobile phone}, [20100601,20100630], [40.25N, 86.54W]>`, in which "msg.ID=201006081123" is automatically generated by the ISP system when accepting the message and attached with the message. The location "$msg.Loc = [40.25N, 86.54W]$" is also automatically attached by ISP system according to the $Macy^*s$ profile.

Definition 3 (Message Request). *A Message Request is a tuple $MsgR$=$< ID, PRE_{issue}, T_{expire}, Loc\_squ >$, where ID is the unique identifier of a user who initiates this request, $PRE_{issue} \subseteq ISSUE$ is a subset of $ISSUE$ denoting the preferred issues, T_{expire} denotes the expired effective date of the requested messages, and Loc_squ is in form of (loc_1, loc_2) denoting the square surrounding the user's position with the top-left point $loc_1 \in POS$ and the bottom-tight point $loc_2 \in POS$.*



(a) Bit mapping coding of categories metadata

(b) The folded coding of an example of user's preference {mobile phone and entertainment} and its storage

Fig. 2. Hierarchical Categories Mapping

An example of message request is `MsgR=<13001234567,{discount,clothes}, 20100701, ([40.25N, 84.54W],[39.25N, 86.27W])>`. This request is sent by the user whose mobile phone number is "13001234567". His interested topics are "discount" and "clothes", and the effective date of messages should not be late than Jun.1 2010. Specially, the location square is generated by the phone application system according to the user's preference (We would discuss this in details in section 5.2).

4 Information Organization and Query Processing

In this section, we would investigate how to organize message information so as to quickly response users' requests. We propose a foldable bit-mapping schema of coding issues for the purpose of efficient storage, which is especially meaningful for mobile applications due to the limit of computing resource and transmission speed. We also present an efficient algorithm to precess user queries.

4.1 An Efficient Coding of Issues and Preferences

The issues considered in a ISP system are hierarchically organized. The categories can be defined by specialists according to semantic structure with the help of classification tool, such as $WordNet$[15]. Each category can be further classified into sub-issues. In our model, the categories of issues and their semantics are called $meta\ data$. For the objective of efficient storage and processing of user query, we introduce the following foldable bit-mapping coding schema:

- Issues are classified into two categories: leaf issue and non-leaf issue. Each issue occupies a bit in a code tuple.
- Each bit is either "0" or "1", respectively denoting issue "unselected" or "selected"
- In a folded code, if the current bit is mapped to a leaf issue, each bit is either 1 or 0; otherwise (namely non-leaf issue) each bit is one of $\{1, 0, *\}$, where "0" (or "1") represents this issue together with its all sub-issues unselected (or selected) and "*" represents some of sub-issues in this category are selected and the details of sub-issue bits are given following this issue bit.

For example, Figure 2 (a) is the given Hierarchical Categories, in which `ISSUE =` $\{$`electric appliance, Sports, Clothes, Entertainment`$\}$,`electric appliance =` $\{$`mobile phone, TV, refrig, Microwave`$\}$, and `Clothes =` $\{$`man, woman, kid`$\}$. Figure 2 (b) shows how to encode a user's preference, say *mobile phone* and *clothes*, into a folded schema "*1000010". It is easy to see that the folded code reduce much storage and thus reduces the transmission time. Similarly, a folded code can be decoded to the unfolded bit-mapping schema under the same principle.

4.2 Message Management and User Query Processing

For the purpose of efficient management, we divide messages into two classes: *effective* and *pending* according to their active time window. A message is *effective* if and only if the current date is within its active time window, while it is *pending* if and only if the current date is ahead of its active time window. A message would be discarded when it expires. Accordingly, two queues are organized for storing messages. A *pending* message is added into the *pending* queue in an ascending order according to the beginning of time window and would be deleted from this queue when entering its active window. The *effective* messages are ordered ascendingly according to the expire date in the *effective* queue and would be deleted when becoming expired.

Generally, there are multiple user requests happened at the same time on ISP server, which could fall into the following modes:

- Category Only Query (CoQ): users only care about the issues of advertisements and the query range is in form of $\perp$. A CoQ query returns advertisements correlative with the preferred issues.
- Location Only Query (LoQ): users only care about geographical range of the messages and the preference field of user request would be set $\emptyset$. A LoQ query returns advertisements within the preferred geographical area.
- Hybrid Query (HQ): is the combination of above two cases.

To efficiently process users requests, we present a fast algorithm to handle users' requests. The main idea on which the algorithm is based is make three comparisons between the users' preferences and messages information. First we compare the issues. Since a user's preferences have been encoded, we need to decode it, as discussed in section 4.1, and store the unfolded code in a variable $IssuesBIT$. If the user query is of LoQ type, each bit of $IssuesBIT$ is set 0. Here we adopt a predicate to perform the logic AND operation on each pair of bits of $IssuesBIT$ and the issues in a message,

Algorithm 1: Handling user queries

```
Require: the *effective* messages queue $ActQ$
    the received user query $UQ$
Ensure: Return the messages $MSG$ that satisfy the conditions in $UQ$
 1: $MSG = \emptyset$
 2: **for** each $MsgR_i \in UR$ **do**
 3:   Get the coded preferred issues $MsgR_i.PRE_{issue}$
 4:   **if** $MsgR_i.PRE_{issue} \neq \emptyset$ **then**
 5:     Unfold $MsgR_i.PRE_{issue}$ into a standard bit mapping $IssuesBIT$
 6:   **else**
 7:     $IssuesBIT = 0$
 8:   **end if**
 9:   $STATE = TRUE$
10:   **while** STATE **do**
11:     sequentially select an element from queue $msg \in ActQ$
12:     **if** $BitLogicAnd(IssuesBIT, msg.Issues)$ AND $Loc\_squ \neq \perp$ **then**
13:       **if** $msg.TW.t_2 \leq MsgR_i.T_{expire}$ **then**
14:         **if** $(MsgR_i.Loc\_squ.lngt_1 \leq msg.loc.long \leq MsgR_i.Loc\_squ.lngt_2)$
            AND $(MsgR_i.Loc\_squ.lat_1 \leq msg.loc.lat \leq MsgR_i.Loc\_squ.lat_2)$
            **then**
15:           $MSG = MSG \cup \{msg.TXT\}$
16:         **end if**
17:       **else**
18:         $STATE = FALSE$
19:       **end if**
20:     **end if**
21:   **end while**
22: **end for**
```

as depicted in step 12. Many programming language provide such function, like C or $C^\sharp$. Then, the algorithm determines whether the expire date of a message is within the user's preferred period, as in step 13. Moreover, if the position of a message's sender is located with the users' preferred geographical range, the message is added to the result set MSG. In special case that users do not care about the geographical position, say $Loc\_squ = \perp$, this check is omitted. Now, we formalize the user request in definition 4 and present the process to handle requests in algorithm 1.

Definition 4 (User Query). *A User Query is a set* $UQ = \{MsgR_1, MsgR_2, \cdots, MsgR_k\}$ *of messages request, where* k *is an integer and each* $MsgR_i, i \in [1..k]$ *is in form of* $MsgR_i$*=*$< ID, PRE_{issue}, T_{expire}, Loc\_squ >$ *including the preferred issues* PRE_{issue}*, the expired effective date* T_{Expire} *and geographical range* Loc_squ*.*

The time computational complexity of algorithm 1 is $O(|UR| * m)$, where $|UR|$ is the number of the message requests in user's query and m is the number of active messages in the IPS store. Since the overwhelming computation is the comparison between requests and messages, the adoption of bit logic operation can highly increase the efficiency.

Fig. 3. The ISP system architecture

Fig. 4. The snapshot of user management of an Intermediary Services Provider system

5 The Prototype System

In this section, we investigate how the proposed model can be realized in the GPS and cellular network systems. We would present the details on design and implementation of the prototype.

5.1 System Architecture of ISP

The architecture of a ISP system is depicted in Figure 3. There are three main components: repositories, internal modules and interfaces. The repositories include the Store of `Registered Merchant Information (RMI)`, the Store of `Registered User Profile (RUI)` and the Store of `Hierarchical Categories (HCS)`. The `RMI` store contains the information associated with the registered merchants, such as geographical information and the certification issued by a government office etc., while `RUI` records users' private information, such as the mobile phone number. `HCS` stores the metadata and the mapping schema of hierarchical categories.

Internal modules, depicted as rectangles in the figure, are the main functions to process merchants advertisements and users queries. The `Identity Management` module is responsible for identifying individuals or merchants and controlling the access to the resources of system by placing restrictions on the established identities. The module of `User Preferences Decoding` deals with the transformation of users' folded preferences into normal bit mapping schema. This module would access the Hierarchical Categories Store `HCS`. The `Categories Maintenance` module provides platform for administrators to create, delete or modify the categories, as well as to maintain the bit mapping schema of categories and storage. The model of `Ad Management` processes the request from registered merchants and manage the *pending* and *effective* message queues. The `Match Preferred Messages` module performs the routine activities of users query on messages that need to access both the `HCS` store and effective messages queue.

The `Merchant Interface` provides the operations to start registration and to complete a publish of an advertisement. The `User Interface` allows users to register on ISP server and perform basic profile management. The `Administrator Interface` offers the management of users, merchants and advertisement. Figure 4 shows the snapshot of user management on our ISP prototype.

5.2 User Location and Communication

Location technologies(LCS) perform the localization of target, and also make the resulting location data available to external actors. There are many methods to location an entity. The basic positioning methods include trilateration, triangulation, hyperbolic. Now

(a) User's preferences (b) The acquired interested messages

Fig. 5. Snapshot of mobile end system

the satellite positioning system has well developed, such as the widely adopted Global Positioning System (GPS) and Galileo. With the development of Mobile-based technologies, positioning in 3G networks are more and more popular, such as the Network-based technologies (TDOA, AOA etc.)

Position appears in the form of spatial coordinates and can be represented as a single point in the Cartesian coordinate, such as $35^0 20' 45.7652'' N - 105^0 26' 9.1432'' W$. In our model, to capture a user's graphical position in real time, we need to associate the point with a certain place in the real world or to map on a descriptive location. We adopt the A-GPS technology to locate a mobile phone user due to its efficient performance. To use the location class `getGps.java`, we need to import two class libraries `javax.microedition.location.Criteria` and `javax.microedition.location.Location`. The adopted classes include `criteria`, `Location` and `Coordinates`. The longitude and latitude of a mobile end can be acquired by following functions `getLocation()`, `LocationProvider.getInstance()`, `getLongitude()` and `getLatitude()`. To protect users' location privacy, we adopt the cloaking technologies. A users is allowed to define a query range $range$. After having the geographical position, say $[x, y]$, where x and y are the coordinates of current point, the square $[x - range, y - range, x + range, y + range]$ is generated and sent to server.

Another two important issues of implementing the prototype are the efficiency and stability of communication between users and server. To avoid the choking consequence caused by wireless transmission, we perform the communication via a separate thread. To increase the stability, we adopt **Hex** code when transmission.

We implement our prototype on a PC with operating system WINDOWS XP SP2, a 2Gz E5200 processor and 2GB RAM. We adopt Tomcat 6.0 as the Web server on ISP and Mysql 5.0 as the database. The development environment on mobil end is Myeclipse 6 and NetBeans 5.51. Figures 5 (a) and (b) are the snapshots on simulated mobile end.

6 Conclusions

Simultaneously supporting personalization and privacy preserving is a challenging problem in mobile marketing. In this paper, we investigate this problem by proposing a Privacy Preserving Model for Personalized Mobile Marketing (P^2PMM). It can protect both users' preferences and location privacy while supporting personalization in mobile marketing. We present an efficient coding method to manage hierarchical categories and users' preferences. We discuss how to apply this model into practice and implement a prototype.

As part of future work, we are planning to investigate the optimization of algorithm so as to efficiently process a large volume of queries. We also plan to address the issue of consideration of more sophisticated constraints.

References

1. MMA: Mobile marketing association: Updates definition of mobile marketing. http://mmaglobal.com/news/mma-updates-definition-mobile-marketing (2009)
2. Marsit, N., Hameurlain, A., Mammeri, Z., Morvan., F.: Query processing in mobile environments: a survey and open problems. In: proceeding of the First International Conference on Distributed Framework for Multimedia Applications. (2005) 150–157
3. Cleff, Beatrix, E.: Privacy issues in mobile advertising. In: British and Irish Law, Education and Technology Association, Annual Conference Hertfordshire. (2007)
4. Hinde, S.: Spam: the evolution of a nuisanceg. Computers and Security **22**(6) (2003) 474C478
5. A, A.S., B, A.D., A, J.M.: Diffusion and success factors of mobile marketing. Electronic Commerce Research and Applications **4**(2) (2005) 159–173
6. RJ, T.I., AG, W.: Consumer responses to interactive advertising campaign coupling short-message-service direct marketing and tv commercials. Journal of Advertising Research **45**(4) (2005) 382–401
7. Kalnis, P., Ghinita, G., Mouratidis, K., Papadias, D.: Preventing location-based identity inference in anonymous spatial queries. IEEE Transactions on Knowledge and Data Engineering **19**(12) (2007) 1719–1733
8. Mokbel, M., Chow, C.Y., Aref, W.: The new casper: query processing for location services without compromising privacy. In: Proceedings of the 32^{nd} International Conference on Very Large Databases (VLDB 2006). (2006)
9. G. Ghinita, M.L. Damiani, E.B.C.S.: Interactive location cloaking with the probe obfuscator. In: In International Conference on Mobile Data Management. (2009)
10. Bertino, E.: Privacy-preserving techniques for location-based services. SIGSPATIAL Special archive **1**(2) (2009) 2–3
11. Cheng, R., Zhang, Y., Bertino, E., Prabhakar, S.: Preserving user location privacy in mobile data management infrastructures. In: In 6^{th} Workshop on Privacy Enhancing Technologies, LNCS 4258. (2006) 393–412
12. Cheng, R., Zhang, Y., Bertino, E., Prabhakar, S.: Preserving user location privacy in mobile data management infrastructures. In: 6^{th} Workshop on Privacy Enhancing Technologies, IEEE (2006) 393–412
13. Watson, R.T., Pitt, L.F., Berthon, P., Zinkhan, G.M.: U-commerce: expanding the universe of marketing. Journal of the Academy of Marketing Science **30**(4) (2002) 333–347

14. Balasubramanian, S., Peterson, R., Jarvenpaa, S.: Exploring the implications of m-commerce for markets and marketing. Journal of the Academy of Marketing Science **30**(4) (2002) 348–361
15. PrincetonUniversity: Wordnet. http://www.wordnet.org/