2014 IEEE 38th Annual International Computers, Software and Applications Conference

Preserving Location-Related Privacy Collaboratively in Geo-Social Networks

Lin Zhang*, Mingxuan Yuan†, Yao Guo*, Xiangqun Chen* and Lei Chen†
* Key Laboratory of High-Confidence Software Technologies(Ministry of Education), Peking Univeristy, China
{zhanglin08, yaoguo, cherry}@sei.pku.edu.cn
†Department of Computer Science and Engineering, Hongkong University of Science and Technology, Hong Kong
{csyuan, leichen}@ust.hk

***Abstract*—The emerging geo-social networks bring us attractive location-based services as well as serious location-related privacy threats. Location information of users in geo-social networks might be revealed by friends carelessly, or deduced by users curiously or even maliciously. In order to avoid location leakages, we propose collaborative privacy management in geo-social networks. Users specify and broadcast their preferences on location-related privacies in advance, so that potential leakages can be reported automatically when new resources arrive. If necessary, the associated spatial and/or temporal information of resources will be tweaked according to the privacy preferences of involving users, so that "old" leakages can be eliminated while ensuring that "new" ones are not introduced. We design algorithms for such tweaks and construct experiments on a simulated dataset to demonstrate their usability and applicability.**

***Keywords*—Collaborative Privacy Management; Location-related Privacy; Geo-social Networks;**

I. Introduction

With the popularity of mobile devices, especially smart phones, resources on social networks are often associated with geo-tags. The resulting geo-social networks provide us attractive location-based services, while introduce serious location-related privacy threats at the same time[1]. For example, one's presence at sensitive places, such as medical institutions and night clubs, might reveal his/her physical situations or personal interests; one's absence from special places, such as home and office, would probably lead to undesirable situations; co-locations of users might be employed to infer their private relationships; and one's sensitive information would not be preserved in case that he/she can be identified determinately in anonymous or pseudonymous geo-social networks.

In many cases, one's locations are revealed in resources posted by his/her careless friends. As a result, collective /collaborative privacy management was proposed and studied [2] [3] [4] [5]. Given a certain resource r, privacy preferences of r's involvers, who are mentioned in r so that sensitive information might be revealed in r, are collected and balanced to make a final decision on r's publication.

However, location leakages still happen when curious or malicious users try to combine resources together to perform inferences. As shown in Figure 1, Alice and Bob are secret lovers, but their co-location in a bar can be inferred from Chris' and David's states so that their private relationship might be further revealed by this co-location with some knowledge background.

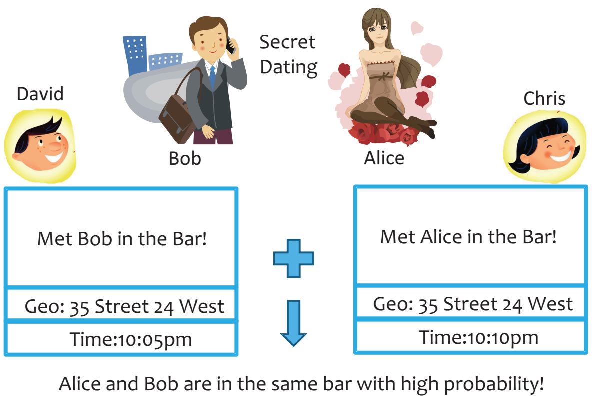

Fig. 1. A Motivating Scenario(Co-Location Inference)

In order to prevent location leakages in geo-social networks, we propose collaborative privacy management. Users specify and broadcast their preferences on location-related privacies in advance, so that potential leakages can be reported and prevented automatically when new resources arrive. The key challenges are listed as follows:

- How to specify location-related privacy preferences?
  In this paper, we introduce the concept of cylinders, which are spatio-temporal regions with additional constraints. Users declare location cylinders, absence cylinders, exclusive cylinders and anonymous cylinders to express their personal preferences on location privacy, absence privacy, co-location privacy and identity privacy, respectively.
- How to report potential location leakages?
  When new resources arrive, existing resources in a particular geo-social network should be employed to infer the presences, absences, co-locations and anonymities of involving users. According to their location-related privacy preferences, violations can be reported automatically.
- How to prevent reported location leakages?
  The associated spatial and/or temporal information of resources should be tweaked accordingly to eliminate re-

0730-3157/14 $31.00 © 2014 IEEE
DOI 10.1109/COMPSAC.2014.94

ported location leakages. More importantly, such tweaks should be careful to remove "old" leakages while ensuring that "new" leakages are not introduced.

II. Privacy Preferences

Users have to specify their personal preferences on location-related privacies in advance. In this paper, we introduce the concept of cylinders, which are spatio-temporal regions with additional constraints:

$$(lat, long, range), (from, to), iteration, constraints$$

where the triple $(lat, long, range)$ specifies a spatial area with the center at $(lat, long)$ and the range of $range$; the bio-tuple $(from, to)$ specifies a temporal interval from $from$ to to. With the triple $(lat, long, range)$ and the bio-tuple $(from, to)$, a spatio-temporal region in three-dimensional can be plotted, and that is why we employ the term of "cylinder" here. $iteration$ describes the rules for repeating the temporal interval, such as daily, workday, offday, weekly, monthly, and so on; $constraints$ represents additional constraints on this cylinder, such as the set of exclusive users on exclusive cylinders, and the threshold on anonymous cylinders.

Location cylinders are used to describe one's location privacy preferences. A location cylinder is specified to express that a particular user does not intend to be located accurately within r_{ij} meters during a time interval from t_i to t_j. Suppose that Alice declares one location cylinder as follows:

$$(*, *, 1000), (0:00, 24:00), daily, *$$

which means that she would not like to be located exactly within one kilometre at any time[1].

Absence cylinders are used to describe one's absence privacy preferences. An absence cylinder is defined to express that a particular user does not intend to reveal his/her absence from a specific place p_{ij} during a time interval from t_i to t_j. Suppose that Alice declares two absence cylinders:

$$(p_{ij}.lat, p_{ij}.long, 1000), (9:00, 12:00), workday, *$$
$$(p_{ij}.lat, p_{ij}.long, 1000), (13:00, 18:00), workday, *$$

where $(p_{ij}.lat, p_{ij}.long)$ represents her office, to express that she does not want to reveal her absence from work during office hours.

Exclusive cylinders are used to describe one's co-location privacy preferences. An exclusive cylinder is specified to express that a particular user does not want to reveal his/her co-location with other users $user_{exclusive}$ during a time interval from t_i to t_j. Suppose that Alice declares her exclusive cylinders as follows:

$$(*, *, 100), (0:00, 8:00), daily, user_{exclusive} = \{Bob\}$$
$$(*, *, 100), (12:00, 13:00), daily, user_{exclusive} = \{Bob\}$$
$$(*, *, 100), (18:00, 24:00), daily, user_{exclusive} = \{Bob\}$$

which means that Alice would like to keep her co-locations with Bob in rest time as secret.

Anonymous cylinders are used to describe one's identity privacy preferences. An anonymous cylinder is specified to express that a particular user does not intend to be identified definitely within anonymous geo-social networks. We employ the concept of k-Anonymity[6] here. Suppose that Alice declares an anonymous cylinder as follows:

$$(*, *, 500), (t-5, t), daily, threshold = 10$$

which means that Alice's identity privacy can be satisfied if and only if she is 10-Anonymous within 500 meters during the past five minutes.

III. Problem Description

Suppose that R is the resource set in a particular geo-social network before r is submitted. We make the following definitions and describe the problem to be solved accordingly.

Given a certain resource r, we employ the following terms, $r.\text{S}$, $r.\text{T}$, and $r.\text{ST}$, to indicate the spatial area occupied by r, the temporal interval occupied by r and the spatio-temporal region occupied by r, respectively.

***Definition** 1:* r is location-privacy preserved if and only if u's location privacy can be preserved for each involving user u of r ($u \in r.\text{WHO}$).

As users' preferences on location privacy are expressed in location cylinders, the necessary and sufficient condition can be formalized as follows:

$$u.Cylinder_{location}\Big|^{lat(D_u),long(D_u)}_{from(D_u),to(D_u)} \subseteq D_u (\forall u \in r.\text{WHO})$$
$$D_u = \bigcap_{r' \in \{r\} \cup R} r'.\text{ST}\Big|^{u \in r'.\text{WHO}}_{r.\text{ST} \cap r'.\text{ST} \neq \emptyset}$$

Simply speaking, other resources in R are combined together with r to deduce u's presence within a more restricted spatio-temporal region D_u for each involving user u of r. We employ the pair of $(lat(D_u), long(D_u))$ to represent D_u's centre point in the spatial plane, and the pair of $(from(D_u), to(D_u))$ to represent D_u's time interval in the temporal axis. We believe that u's preferences on location privacy are satisfied if and only if D_u is more general than the one specified in u's location cylinders.

***Definition** 2:* r is absence-privacy preserved if and only if u's absence privacy can be preserved for each involving user u of r ($u \in r.\text{WHO}$).

As users' preferences on absence privacy are expressed with absence cylinders, the necessary and sufficient condition can be formalized as follows:

$$u.Cylinder_{absence}\Big|_{from(D_u),to(D_u)} \cap D_u \neq \emptyset (\forall u \in r.\text{WHO})$$
$$D_u = \bigcap_{r' \in \{r\} \cup R} r'.\text{ST}\Big|^{u \in r'.\text{WHO}}_{r.\text{ST} \cap r'.\text{ST} \neq \emptyset}$$

***Definition** 3:* r is co-location privacy preserved if and only if u's co-location privacy can be preserved for each involving user u of r ($u \in r.\text{WHO}$).

[1] $*$ is the wildcard, indicating that no explicit constraints are set here.

As users' preferences on co-location privacy are expressed with exclusive cylinders, the necessary and sufficient condition can be formalized as follows:

$$r.\text{ST} \cap r'.\text{ST}\big|_{r'.\text{WHO} \cap E_u \neq \emptyset}^{r' \in \{r\} \cup R} = \emptyset$$
$$E_u = \{u' | u' \in u.Cylinder_{exclusive}.user_{exclusive}\}$$

Simply speaking, the set of exclusive users E_u is calculated for each involving user u of r. And we believe that u's preferences on co-location privacy are satisfied if and only if the co-location of u and u's exclusive users can not be inferred from resources in $\{r\} \cup R$.

Definition 4: r is identity privacy preserved if and only if u's identity privacy can be preserved for each involving user u of r ($u \in r.\text{WHO}$).

As users' preferences on identity privacy are expressed with anonymous cylinders, the necessary and sufficient condition can be formalized as follows:

$$u.Cylinder_{anonymous}.threshold \leq |A_r|$$
$$A_r = \bigcup_{\substack{r' \in \{r\} \cup R \\ r'.\text{ST} \cap u.Cylinder_{anonymous}\big|_{from(r.\text{ST})}^{lat(r.\text{ST}, long(r.\text{ST}))} \neq \emptyset}} r'.\text{WHO}$$

Simply speaking, the set of nearby resources A_r is calculated accordingly. And we believe that u's preferences on identity privacy are satisfied if and only if the number of nearby users are more than the $threshold$ in u's anonymous cylinders.

Problem Description 1: In order to protect the location-related privacy of users in a real-name geo-social network, a certain r is allowed to be shared if and only if the following requirements can be satisfied simultaneously:

- r is location-privacy preserved.
- r is absence-privacy preserved.
- r is co-location-privacy preserved.

Otherwise, r's associated spatial and/or temporal information should be tweaked until these requirements are achieved.

Problem Description 2: In order to protect the location-related privacy of users in an anonymous geo-social network, a certain resource r is allowed to be shared if and only if r is identity-privacy preserved. Otherwise, r's associated spatial and/or temporal information will be tweaked until the above requirement is achieved.

IV. Tweak Algorithms

Based on definitions in Section III, it is quite straightforward to report potential location leakages. In this section, we focus on tweaking the associated spatial and/or temporal information of resources, so that "old" leakages can be eliminated while "new" ones cannot be introduced.

Since resources in geo-social networks might arrive in bathes with high probabilities, we suppose that n resources are simultaneously submitted by different users, which can be referred as G. For each resource $r_i (1 \leq i \leq n)$ in G, the corresponding $r_i.\text{U}$ and $r_i.\text{R}$ are calculated as follows:

$$r_i.\text{U} = r_i.\text{WHO} \cup \{u' | u' \in E_u (u \in r_i.\text{WHO})\}$$
$$r_i.\text{R} = \{r' | r'.\text{WHO} \cap r_i.\text{U} \neq \emptyset, r'.\text{ST} \cap r.\text{ST} \neq \emptyset (r' \in R \cup G)\}$$

We regard each resource as a vertex, and connect two vertices v_i and v_j with an edge if $r_i.\text{U} \cap r_j.\text{U}$ is not empty. On the resulting graph, we employ the depth first search algorithm (DFS) to split n resources into subgroups.

For each subgroup g, if potential location-related privacy leakages can be reported, we employ a generic algorithm to re-arrange the resources in g within the following spatio-temporal region $g.\text{D}$:

$$g.\text{D}.lat_{min} = \min\{r'.\text{S}.lat_{min} | r' \in \cup_{r_i \in g}(r_i \cup r_i.\text{R})\}$$
$$g.\text{D}.lat_{max} = \max\{r'.\text{S}.lat_{max} | r' \in \cup_{r_i \in g}(r_i \cup r_i.\text{R})\}$$
$$g.\text{D}.long_{min} = \min\{r'.\text{S}.long_{min} | r' \in \cup_{r_i \in g}(r_i \cup r_i.\text{R})\}$$
$$g.\text{D}.long_{max} = \max\{r'.\text{S}.long_{max} | r' \in \cup_{r_i \in g}(r_i \cup r_i.\text{R})\}$$
$$g.\text{D}.t_{from} = \min\{r'.\text{T}.from | r' \in \cup_{r_i \in g}(r_i \cup r_i.\text{R})\}$$
$$g.\text{D}.t_{to} = \max\{r'.\text{T}.to | r' \in \cup_{r_i \in g}(r_i \cup r_i.\text{R})\}$$

A. Generic Algorithm

Suppose that there are m resources in g. An m-length sequence of candidates $(d_i.lat, d_i.long, d_i.t)(1 \leq i \leq m)$ can be regarded as an arrangement of resources in g. We then encode it into a $3 \times 16 \times m$-length (0,1)-string as an individual in the population.

The algorithm starts with ten random individuals, and employs the $\mathbf{KS}(id)$ as the fitness function, which indicates the number of location-related privacy-preserved resources in $g.\text{R}$:

$$\mathbf{KS}(id) = KS(d_1, d_2, ..., d_m)$$

A population generation is evolved are as follows:

- Selection
  We calculate $\mathbf{KS}(id)$ for each individual in the population, and then associate each individual with the probability $p(id)$:
  $$p(id) = \frac{\mathbf{KS}(id)}{\sum_{i=1}^{10} \mathbf{KS}(id)}$$
  Individuals are then randomly selected with roulette.
- Crossover
  Ten individual are classified into five pairs randomly. For each pair of individuals, we generate a random index $rand$ $(0 < rand < 3 \times 16 \times m)$, and then employ single-point crossover.
- Mutation
  For each individual, we generate a new random index $rand$ $(0 < rand < 3 \times 16 \times n)$, and then employ the single-point mutation.
- Marker Selection
  Individuals, which are processed with selection, crossover and mutation, are regarded as the next generation. We select the individual id with the highest $\mathbf{KS}(id)$ as the marker for the coming round(s).

The algorithm will terminate if one hundred generations have been produced, or the value of function $KS()$ reaches $|\cup_{r_i \in g} r_i.\text{R}| + m$.

Algorithm 1: Annealing Algorithm

```
Input:
A certain resource $r$
Output:
A candidate $d$ for $r$'s mock spatio-temporal region
1  set $T_{max} = 10$;
2  set $T_{min} = 0.1$;
3  set $\rho = 0.95$;
4  set $T = T_{max}$;
5  set $d_{lat} = r.\text{D}.lat_{max} - r.\text{D}.lat_{min}$;
6  set $d_{long} = r.\text{D}.long_{max} - r.\text{D}.long_{min}$;
7  set $d_t = r.\text{D}.t_{to} - r.\text{D}.t_{from}$;
8  set $d = (r.\text{D}.lat_{min} + \frac{d_{lat}}{2}, r.\text{D}.long_{min} + \frac{d_{long}}{2}, r.\text{D}.t_{from} + \frac{d_t}{2})$;
9  set $s_d = \text{KS}(d)$;
10 while $T \geq T_{min}$ do
11     set $lat = r.\text{D}.lat_{min} + rand() \times d_{lat}$;
12     set $long = r.\text{D}.long_{min} + rand() \times d_{long}$;
13     set $t = r.\text{D}.t_{from} + rand() \times d_t$;
14     set $d' = (lat, long, t)$;
15     set $s_{d'} = \text{KS}(d')$;
16     set $dS = s_{d'} - s_d$;
17     if $dS > 0$ then
18         set $d = d'$, $s_d = s_{d'}$;
19     else if $\exp(dS/T) > rand()$ then
20         set $d = d'$, $s_d = s_{d'}$;
21     set $T = \rho \times T$;
```

B. Annealing Algorithm

Suppose that, there are only one resource r in g. We calculate $r.\text{U}$ and $r.\text{R}$ at first, and then employ the annealing algorithm 1 to generate a mock spatio-temporal region for r within the following spatio-temporal region $r.\text{D}$:

$$
\begin{aligned}
r.\text{D}.lat_{min} &= \min\{r'.\text{S}.lat_{min} | r' \in r \cup r.\text{R}\} \\
r.\text{D}.lat_{max} &= \max\{r'.\text{S}.lat_{max} | r' \in r \cup r.\text{R}\} \\
r.\text{D}.long_{min} &= \min\{r'.\text{S}.long_{min} | r' \in r \cup r.\text{R}\} \\
r.\text{D}.long_{max} &= \max\{r'.\text{S}.long_{max} | r' \in r \cup r.\text{R}\} \\
r.\text{D}.t_{from} &= \min\{r'.\text{T}.from | r' \in r \cup r.\text{R}\} \\
r.\text{D}.t_{to} &= \max\{r'.\text{T}.to | r' \in r \cup r.\text{R}\}
\end{aligned}
$$

The centre of $r.\text{D}$ is assumed as the initial candidate. The function $KS()$, which returns the number of location-related privacy-preserved resources in $r.R$, is employed as the energy function.

We initialize controlling variables from Line 1 to Line 4, constants from Line 5 to Line 7, and state variables from Line 8 to Line 9. Before the annealing process terminates, a candidate d' is randomly generated within $r.\text{D}$. If d' is a more suitable candidate (From Line 17 to Line 18), we regard d' as the "marker" for the coming round(s); Otherwise, d' will be dropped unless its accepted probability is high enough (From Line 19 to Line 21).

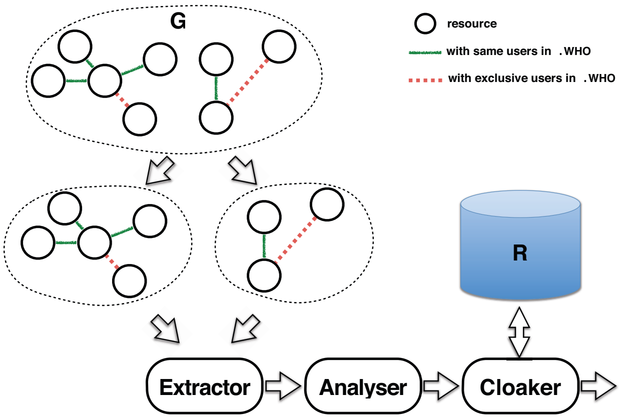

Fig. 2. Framework

V. Prototype Design

As shown in Figure 2, when new resources arrive, they are split into subgroups at first, and then processed independently.

The characteristics of a specific resource r are extracted automatically, to decide the set of involving users r.WHO, the occupied spatial area r.S, the occupied temporal interval r.T and the occupied spatio-temporal region r.ST.

A. Extracting WHO

Given a certain resource r, we employ the word segmentation service provided by Sina to pre-process r at first. In this way, the temporal and spatial expressions in r will be labelled and picked out. We assume each left word segment as a candidate name for further process, so that resource co-owners will not be left out carelessly.

Taking nicknames and spelling mistakes into considerations, we employ a fuzzy name matching algorithm [7] for the WHEN extraction, and set the threshold for similarity as 0.85. It means that we regard the two names as the same if the overlapping ratio of these two candidates is equal to or higher than 0.85.

B. Extracting WHEN

A public temporal dictionary is established in advance with temporal words and phrases. Based on which, the time stamp of a resource r is regarded as the temporal start-point, and the temporal expressions in r are detected and adjusted accordingly.

For example, "Alice 2012-02-12 13:45:53 Lucky! Two tickets for Wallace's concert next Saturday night!". Two temporal expressions can be segmented and labelled, which are "next Saturday" and "night". Since the time stamp of this resource is "2012-02-12 13:45:53", its WHEN characteristic can be adjusted to "2012-02-18 21:00:00".

C. *Extracting* WHERE

A public spatial dictionary is established in advance with spatial words and phrases. Based on which, the geo-tag of a resource r is regarded as the spatial region, and the spatial expressions in r are detected and adjusted accordingly.

Suppose that the former status is associated with the geo-tag of "Shanghai". One spatial expression "Wallace's concert" can be segmented and labelled. As a result, its WHERE characteristic can be adjust to "31.206324 121.459388"(Shanghai Luwan Gymnasium) with the help of Google Map service.

VI. Experiment

Experiments are constructed to evaluate the proposed algorithms. Since human efforts are needed during evaluation, we built our experiments on a simulated dataset instead of a real one.

A. *Simulation*

MilanNight[8] is a simulated dataset of user movement from 7:00 PM to 1:00 AM in the city of Milan. Since friendships between users are unavailable in MilanNight, we employ the R-MAT graph mode[9] to generate a simulated social network. As demonstrated in [9], we set parameters a, b, c and d as 0,45, 0.15, 0.15 and 0.25, respectively.

With the simulated social network, we query for uid's friends within 20 meters to generate a resource for each record $(uid, timestamp, longitude, latitude)$ in MilanNight. Its content is randomly selected from 200 templates, which are provided by 15 volunteers in our department. Finally, we got 35,618 resources in the total.

Privacy cylinders of users are randomly generated and attached. For each user uid in MilanNight, the number of his/her cylinders varies from 3 to 5.

B. *Accuracy*

We study the precision and recall of characteristic extraction on WHEN, WHERE, and WHO, respectively.

1) WHO *Extraction:* Among the 200 resources provided by our volunteers, there are 47 resources containing explicit user names or nicknames. Our prototype application reports 61 resources, out of which 43 are correctly extracted. Consequently, the precision of the "WHO" information extraction is $43/61 = 70.48\%$, and the recall of the "WHO" information extraction is $43/47 = 91.49\%$. In the 47 messages containing explicit user names or nicknames, there are 57 names in the total. Our prototype application reports 77 names, out of which 53 are correct. It means that the precision of name/nickname extraction is $53/77 = 68.83\%$, and the corresponding recall is $53/57 = 92.98\%$.

2) WHEN *Extraction:* Since we take the time stamp of a resource as default when there are not explicit temporal expressions, the recall of the WHEN characteristic extraction in our prototype is always 100%. As the result, we just evaluate the accuracy of the WHEN extraction with precision. Among the 200 resources provided, there are 179 resources whose WHEN information can be extracted correctly, which means that the precision of the WHEN information extraction is $179/200 = 89.50\%$.

TABLE I
Average Cost for Candidate Generation(s)

(a) Single-resource Subgroup

Window Size	r.R				
	<10	10-15	15-20	20-25	>30
00:30:00	3.53	4.17	4.20	6.08	6.74
01:00:00	4.23	5.28	5.14	7.89	8.74
01:30:00	4.69	6.94	6.85	9.99	14.48
02:00:00	5.11	8.33	7.37	12.00	14.87
02:30:00	5.69	8.48	7.98	12.02	15.96
03:00:00	6.26	9.16	8.58	15.98	21.06
03:30:00	7.96	12.39	13.24	20.45	22.64
04:00:00	8.16	14.15	14.04	24.26	25.79
04:30:00	8.19	13.87	14.37	22.56	24.33
05:00:00	8.34	14.42	14.04	23.02	25.67
05:30:00	8.44	15.51	16.22	24.54	26.96
06:00:00	8.64	15.54	16.05	24.09	26.13

(b) Double-resource Subgroup

Window Size	g.R				
	<10	10-15	15-20	20-25	>30
00:30:00	8.06	10.23	12.20	16.44	22.10
01:00:00	9.17	11.59	14.03	17.28	24.02
01:30:00	10.14	13.62	15.97	19.52	27.09
02:00:00	10.63	15.22	17.33	20.04	28.63
02:30:00	12.27	16.32	16.64	22.11	31.30
03:00:00	13.63	18.24	19.07	23.92	33.67
03:30:00	15.02	19.17	22.43	25.06	34.18
04:00:00	15.99	19.82	23.84	27.00	36.29
04:30:00	17.32	21.23	25.08	28.93	37.24
05:00:00	17.93	22.17	24.92	29.01	38.62
05:30:00	18.44	23.68	27.96	31.02	40.17
06:00:00	19.62	23.59	29.03	33.19	42.48

3) WHERE *Extraction:* Resources in geo-social networks always carry geo-tags, so that the recall of the WHERE characteristic extraction is always 100% even though explicit spacial expressions can not be detected. As the result, we will just evaluate the accuracy of the WHERE characteristic extraction with precision. Among the 200 resources provided, there are 163 resources whose WHERE characteristic can be extracted correctly. The precision is $163/200 = 81.50\%$.

C. *Performance*

We run our algorithms on a computer with the 24 Intel(R) Xeon(R) CPU X5650 @2.67GHz and 4 × 4 GB DDR3 Memory @1333MHz.

In Section III, potential location-related leakages are reported with existing resources in a particular geo-social network. However, it will be time-consuming and heavy-computing if the entire set R are always considered when new resources come. We assume resources issued 0.5h, 1h, 1.5h, ..., 6h earlier as R, respectively. In our experiments, small subgroups (single-resource subgroups, double-resource subgroups, and four-resource subgroups) are very common. Due to space limitation, we only demonstrate the average costs for candidate generation with single-resource subgroups and double-resource subgroups here, as shown in Table I.

VII. Related Work

Researches about privacy in social networks can be generally divided into three categories. Since current privacy settings in social networks are too complicated for average users to handle, and too time-consuming for advanced users to configure[10], assistant tools are proposed and developed [11] [12] [13] [14]. Since sensitive information of users can be inferred with majority voting, community detection or classification techniques, researches are proposed to import anonymous principles, such as k-anonymity[6], l-diversity[15], t-closeness[16] and differential privacy, into social networks. Since multi-party authorities of resources often bring privacy conflicts, new access control models are proposed and studied[2] [3] [4] [5].

Techniques, which are widely used to address privacy threats in location-based services, can also be employed in geo-social networks, such as query enlargement[17], fake locations[18], and encryption techniques[19].

Privacy issues in geo-social networks focus on the protection of user locations (both historical locations and current ones), which are quite different from that in social networks. In [1], four typical privacy threats in geo-social networks are summarized, including location privacy, absence privacy, co-location privacy and identity privacy. In [3], the authors studied how to preserve location and absence privacy in geo-social networks. And to the best of our knowledge, there is only one research about co-location privacy in geo-social networks[20].

VIII. Conclusion

Resources in geo-social networks, especially the ones issued from the mobile, are usually associated with geo-tags. As a result, location-related privacies of users are suffering from resources posted by careless friends and inferences performed by curious/malicious users.

In this paper, we proposed collaborative privacy management, in which users specify their location-related privacy preferences in advance, so that potential location leakages can be prevented automatically through the collaboration between users. We introduced the concept of cylinders for users to express their preferences on different location-related privacies, including location privacy, absence privacy, co-location privacy and identity privacy. We formulated the definitions of location-related privacy-preserved resources and described the problem accordingly. Generic and annealing algorithms are designed to tweak the associated spatial and/or temporal information when necessary, making "old" potential leakages eliminated while preventing "new" ones from arising. Experiments are then constructed on a simulated dataset to demonstrate their usability and applicability.

Acknowledgement

This work is supported in part by the High-Tech Research and Development Program of China under Grant No. 2013AA01A605, the National Basic Research Program of China (973) under Grant No. 2011CB302604, the National Natural Science Foundation of China under Grant No. 61103026, 61121063, U1201252, and the NSFC/RGC Joint Research Project (No.60931160444).

References

[1] C. Ruiz Vicente, D. Freni, C. Bettini, and C. Jensen, "Location-Related Privacy in Geo-Social Networks," *Internet Computing*, vol. 15, no. 3, pp. 20 –27, may-june 2011.

[2] A. C. Squicciarini, M. Shehab, and F. Paci, "Collective Privacy Management in Social Networks," in *Proceedings of the 18th international conference on World Wide Web*, 2009, pp. 521–530.

[3] D. Freni, C. Ruiz Vicente, S. Mascetti, C. Bettini, and C. S. Jensen, "Preserving Location and Absence Privacy in Geo-Social Networks," in *Proceedings of the 19th ACM international Conference on Information and Knowledge Management*, 2010, pp. 309–318.

[4] H. Hu, G.-J. Ahn, and J. Jorgensen, "Detecting and Resolving Privacy Conflicts for Collaborative Data Sharing in Online Social Networks," in *Proceedings of the 27th Annual Computer Security Applications Conference*, 2011, pp. 103–112.

[5] H. Hu and G.-J. Ahn, "Multiparty Authorization Framework for Data Sharing in Online Social Networks," in *Proceedings of the 25th annual IFIP WG 11.3 conference on Data and Applications Security and Privacy*, 2011, pp. 29–43.

[6] L. Sweeney, "k-Anonymity: a Model for Protecting Privacy," *International Journal of Uncertainty Fuzziness Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, Oct. 2002.

[7] J. Vosecky, D. Hong, and V. Y. Shen, "User Identification across Social Networks using the Web Profile and Friend Network," *International Journal of Web Applications*, vol. 2, pp. 23–34, 2010.

[8] S. Mascetti, D. Freni, C. Bettini, X. S. Wang, and S. Jajodia, "On the Impact of User Movement Simulations in the Evaluation of LBS Privacy Preserving Techniques," in *European Symposium on Research in Computer Security*, 2008.

[9] D. Chakrabarti, Y. Zhan, and C. Faloutsos, "R-MAT: A Recursive Model for Graph Mining," in *SDM*, 2004.

[10] R. Gross and A. Acquisti, "Information Revelation and Privacy in Online Social Networks," in *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, 2005, pp. 71–80.

[11] L. Fang and K. LeFevre, "Privacy Wizards for Social Networking Sites," in *Proceedings of the 19th international conference on World Wide Web*, 2010, pp. 351–360.

[12] G. Danezis, "Inferring Privacy Policies for Social Networking Services," in *Proceedings of the 2nd ACM workshop on Security and artificial intelligence*, 2009, pp. 5–10.

[13] H. R. Lipford, A. Besmer, and J. Watson, "Understanding Privacy Settings in Facebook with an Audience View," in *Proceeding of the 1st Conference on Usability, Psychology, and Security*, 2008, pp. 2:1–2:8.

[14] K. Liu and E. Terzi, "A Framework for Computing the Privacy Scores of Users in Online Social Networks," *ACM Transactions on Knowledge Discovery from Data*, vol. 5, no. 1, pp. 6:1–6:30, Dec 2010.

[15] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-Diversity: Privacy Beyond k-Anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, Mar. 2007.

[16] N. Li and T. Li, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," in *In Proceeding of IEEE 23rd International Conference on Data Engineering*, 2007.

[17] S. Mascetti, D. Freni, C. Bettini, X. S. Wang, and S. Jajodia, "Privacy in Geo-social Networks: Proximity Notification with Untrusted Service Providers and Curious Buddies," *The VLDB Journal*, vol. 20, no. 4, pp. 541–566, Aug. 2011.

[18] H. Kido, Y. Yanagisawa, and T. Satoh, "An Anonymous Communication Technique using Dummies for Location-Based Services," in *Proceedings of International Conference on Pervasive Services*, 2005, pp. 88 – 97.

[19] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private Queries in Location Based Services: Anonymizers are not Necessary," in *Proceedings of the 2008 ACM SIGMOD international conference on Management of Data*, 2008, pp. 121–132.

[20] M. Camilli, "Preserving Co-Location Privacy in Geo-Social Networks," *in press*, 2012.