SALS: Semantics-Aware Location Sharing Based on Cloaking Zone in Mobile Social Networks

Yanzhe Che
College of Computer Science
and Technology
Zhejiang University, Hangzhou
PR China, 310027
pomme@zju.edu.cn

Kevin Chiew
School of Engineering
Tan Tao University
Duc Hoa District
Long An Province, Vietnam
kevin.chiew@ttu.edu.vn

Xiaoyan Hong
Department of Computer
Science
University of Alabama
Tuscaloosa, AL 35487, USA
hxy@cs.ua.edu

Qinming He
College of Computer Science
and Technology
Zhejiang University, Hangzhou
PR China, 310027
hqm@zju.edu.cn

ABSTRACT

There is a potential privacy breach when users access various location-based social applications on a mobile social network (MSN), e.g., sharing locations with friends. To preserve location privacy, one of the most common methods is to use a coarse or fake location instead of a user's exact location. However, most of these previous approaches only provide geometric strategies without considering the semantic context of the geographical locations. For example, if a cloaked region contains a part of a lake, where no boats are allowed, an adversary can easily prune the cloaked region to a smaller range covering a user's actual location.

In this paper, we propose SALS, a semantics-aware location sharing framework based on cloaking zone for an MSN environment. By considering a user's social relations and activities which are available in an MSN environment, SALS does not assume any trustworthy entities, including strangers, friends or any third parties. As a solution, SALS enables users to cooperate with each other, in a Peer-to-Peer (P2P) way, to generate the cloaking zones, which will be used instead of the actual locations. Different from the previous cloaking techniques, SALS considers the semantic location which can influence the distribution probability of a user's locations. We also propose metrics for measuring the quality of the cloaking zone. The evaluation shows that our method can well defend the *semantic-location attack*.

Categories and Subject Descriptors

K.4 [**Computers and Society**]: Public Policy Issues—*Privacy*; C.2.0 [**Computer-Communication Networks**]: General—*Security and protection*; H.2.8 [**Information system**]: Database Applications—*Spatial databases and GIS*

General Terms

Security, Management

Keywords

Mobile social network, Location privacy, Semantic location, Cloaking zone

ACM SIGSPATIAL MobiGIS'12, November 6, 2012. Redondo Beach, CA, USA
Copyright (c) 2012 ACM ISBN 978-1-4503-1699-6/12/11 ...$15.00.

1. INTRODUCTION

The rapid evolution of the mobile devices has enabled the conventional web-based social network, such as Facebook, with ubiquitous accessibility and location-based feature which are turning it mobile. On the other hand, native mobile social networks (MSNs), such as Foursquare, are created dedicated to mobile users which can establish mobile virtual communities to bring individuals with similar interests together via their smartphones or tablets. As a result, both of the two trends have motivated the great popularity of MSNs nowadays. Embedded with location-based features, MSNs are capable of providing diverse location-based social applications, e.g., "check in"[1], personal life sharing[2], location-based social games[3], as well as some traditional location-based services (LBS) like restaurant recommendations. While using these applications, mobile users need to report their locations to the MSN servers timely or to share locations with their social friends or some strangers.

However, the widespread use of location information also raises a great concern for preserving location privacy. On one hand, some advanced location-determining techniques, e.g., GPS and RFID, can precisely determine one's position both indoor and outdoor with an accuracy of about 10 meters which has made it easier for an adversary to disclose one's private information. For example, by inferring from

[1] Provided by Foursquare at `https://foursquare.com`
[2] Provided by Path at `https://path.com`
[3] Shadow Cities at `http://www.shadowcities.com`

a user's whereabouts, an adversary can precisely determine the user's home address. On the other hand, a user's location data, if combined with his profile and social information which are available on an MSN, may cause even more serious threats. For example, if a mobile user, whose identity is available on an MSN, is found to visit a cancer hospital by his insurance agent, he may have to pay more on his medical insurance in the next year.

In order to address such problems, some of the previous work proposed solutions of both pseudonyms and dummy location sharing [1, 2, 13]. Other research [8, 12, 15] designed policies to control the sharing process as well as the usage of these location information. We also have many existing spatial cloaking techniques to solve this problem [3,5,6,9,10,14], which will hide users inside a coarse area to achieve the k-anonymity protection. A common drawback for all of the above solutions is that they all assume some trustworthy entities in their frameworks, e.g., a third party server or some neighbor mobile users. However, in an MSN environment, none of these entities can be fully trusted, e.g., the location server which stores users' actual locations, a third party server which manages the dummy-ID pool for users or the social friends on MSN who may accidentally misuse the users' shared locations. Another serious problem is, although there already exist solutions that can defend various attacks, e.g., *social-relation-disclosure attack* [2], and solutions which provide different geometric strategies to achieve good k-anonymity protection, however most of them do not consider the background knowledge of the geographical context, i.e., the semantic location. A semantic location is a geographical place with real-life context, e.g., a supermarket or a lake, which can affect the distribution probability of users' locations. If an adversary has such knowledge, he can prune a user's coarse location into a precise bound and thus defeat these privacy-preserving solutions.

Our research takes full account of the different types of privacy attacks in an MSN environment and proposes a framework of exhaustive location privacy preserving for mobile users, namely SALS. In SALS, we do not trust any entities in an MSN environment, thus every user never reports the exact location but uses a customized cloaking zone (CZ) instead. The CZ covers the user himself as well as some anonymities and achieves the k-anonymity protection. While generating a CZ, SALS enables mobile users to work together, in a P2P way, to make the CZ flexible enough to cover more anonymities inside. Moreover, SALS also considers the semantic location affection and hence adjusts the shape and area of the CZ in order to avoid covering some unreachable places. For example, to cover a lake as one part of the CZ is meaningless because few mobile users appear on the lake generally. In brief, the major contributions of our work are summarized as below.

- We propose a privacy preserving framework SALS for an MSN environment without assuming any trustworthy entities.

- We summarize two main categories of privacy attacks and describe the main idea for each attack.

- When generating a CZ, we take into account the semantic locations affection. We also specify the metrics for measuring the quality of a CZ and propose *MaxDen* algorithm for generating a CZ.

- Experimental results show that our *MaxDen* algorithm has a good performance to defend the *semantic-location attack*.

The remaining sections are organized as follows. After reviewing the related work in Section 2, we analyze the potential threats and summarize the privacy attacks in an MSN environment in Section 3, and propose our research problem and the SALS architecture in Section 4, followed by presenting the design of SALS framework in Section 5. We give the implementation of *MaxDen* algorithm for generating CZ in Section 6 with experiments to evaluate the performance against privacy attacks in Section 7 before concluding the paper in Section 8.

2. RELATED WORK

For a broad range of techniques proposed for preserving user locations privacy, we categorize them into three main classes, i.e., *pseudonyms and dummy location technique*, *privacy policies technique*, and *spatial cloaking technique*.

As a representative of the pseudonyms and dummy location technique, mix-zone [1] enables users to change their pseudonyms inside a special region where users do not report locations. However, it is not suitable for an MSN because an MSN server has to maintain the right social relation for each user, making the pseudonyms useless. On the other hand, some work [2, 13] enables users to report dummy locations to location servers. However, it brings with both extra computing and communicating overheads to generate, transmit and process these dummy locations.

Some privacy-policies solutions design several policies, e.g., strict access control rules and location usage strategies, to protect users' privacy when sharing the location information [8, 12, 15]. The IETF geopriv working group[4] also provided several protocols and frameworks for the enforcement of location privacy policies.

Many solutions employ spatial cloaking technique for location privacy preserving in which users are hidden inside a cloaking region to achieve the k-anonymity protection which means a user can hide inside $k-1$ anonymities from which an attacker cannot distinguish him/her. Some of them are built on a trust third party architecture [9, 10, 14], whilst some other of them use a P2P way [3, 5, 6]. All of these solutions rely on some trust relationships, either the third parties or other mobile P2P users. However in an MSN environment, these entities, including the MSN servers as well as a user's social relationships, are not always trustworthy. Moreover, most of these solutions do not consider the semantic location affection. In this paper, our solution follows this category of technique and tries to solve the trust problem with the consideration of semantic location affection.

3. THREATS AND ATTACK MODELS

In this section, we present the main location privacy threats and then describe different types of attacks that an adversary can use to violate a user's location privacy.

3.1 Potential Threats

Based on the unique characteristic of an MSN environment, we analyze the potential threats for mobile users.

[4] Geopriv `http://datatracker.ietf.org/wg/geopriv/charter`

Threat 1. *The mobile users cannot be trusted.* One of the most important characteristics of an MSN is that we cannot trust a mobile user only based on his online identity. This is because there are some malicious attackers who secretly collecte a victim's location information or actively attack the victim. Such an attacker can use a pseudonyms/fake online identity to pretend to be a normal user or can compromise some of the victim's friends' mobile devices. As a result, we conclude that the mobile users, no matter strangers or friends, should never be fully trusted for privacy concern.

Threat 2. *The MSN servers or other third parties cannot be trusted.* The fact is that the MSN servers and other third parties, who provide location-based social applications, have already record a list of historical location information of each user. By utilizing various types of data mining techniques, they can disclose users' interests and habits, or trace their trajectories. It is not surprising to see that they trade the users' valuable personal information for commercial purpose or with political motivation. Moreover, the MSN servers could be compromised by an adversary, leading to further disclosure of the sensitive information for malicious purpose.

3.2 Attack Models in an MSN Environment

This section summarizes the attack models in an MSN environment where an adversary can collect data from the MSN applications to violate a user's location privacy. In this paper, we classify all the attacks into two main categories, namely *passive attacks* and *active attacks.*

For the *passive attacks*, an adversary passively collects mobile users' location information and estimates a user's location by (1) utilizing the flaw of the user's location generating algorithm; (2) utilizing some background knowledge of the geographical context, i.e., semantic locations; or (3) utilizing the social relationship information of the user. A group of attacks for this category are given as follows.

In the *semantic-location attack*, based on the knowledge of the geographical location around a victim, an adversary can take advantage of the semantic location affection to prune the victim's coarse location into a precise bound [7]. For example, the victim creates a cloaking region which covers a shopping mall inside. However, a malicious attacker, who is familiar with this area, knows that the shopping mall is closed at this time so that he can prune the cloaked region to estimate the victim's real location.

Due to different strategies of cloaked region generating algorithms, a victim may have a higher probability to be located around the center or the boundary of the cloaked region. Therefore, a *center-of-zone attack* or a *border-of-zone attack* can be used to guess the victim's exact location in the given region.

In an MSN environment, where users' social relationship is available, an adversary can also initiate a *social-relation-disclosure attack* in which a victim's real location may be disclosed by his online friends. For example, when visiting a cancer hospital, Alice stops sharing location information for privacy concern. However, her accompanying friend Bob is still sharing locations. Therefore, an adversary who is aware of the friendship can also disclose Alice's activity illegitimately.

For the *active attacks*, an adversary not only collects but also actively sends malicious messages to a victim in order to disturb his normal privacy-preserving process. In the *region-probing attack*, an adversary may send several fake locations, which are purposely created, to the victim. Suppose that the victim has adjusted his cloaked region in order to cover these fake locations, then the adversary can estimate the victim's location based on the change of the cloaked region.

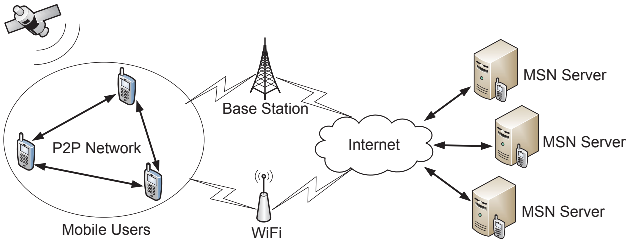


Figure 1: The system architecture of SALS

In the *location-flooding attack*, which is similar to the Denial-of-Service (DoS) attack, a victim will receive flooding dummy locations, which are sent by an attacker or forwarded by some innocent users, so that the mobile device cannot work normally. We also introduce the *visual-aid attack* in which an adversary can observe the locality directly or via some security cameras. For example, a user has reported many fake locations around him. However the cameras show that the user is the only person inside that locality and then an attacker can identify him. Furthermore, the adversary may also compromise the MSN servers or mobile devices directly. However, due to the page limitation, we leave addressing these active attacks to future work.

4. SYSTEM MODEL AND PROBLEM STATEMENT

In this section, we present our research problem and then describe the system architecture of SALS.

4.1 Research Problem

Given all these threats and attacks leading to privacy violation, the research problem of this paper is to design an efficient and effective framework to enable users to cooperate together to generate customized cloaking zones which are used for location sharing instead of their actual locations.

Note that the framework should also satisfy the following requirements. (1) A user does not have to trust any entities in an MSN environment. (2) The generated CZ should try to cover more anonymous users inside to achieve better k-anonymity protection. (3) The generated CZ should also consider semantic location affection.

4.2 System Architecture

Different from some of the former work such as MobiShare [13], we do not deploy a location server to store users' location information. Instead, the SALS architecture is similar to the conventional client-server architecture which is simple and easy to implement. Figure 1 shows the details of the architecture. It only consists of two types of entities, namely mobile users and MSN Servers, and two types of networks, i.e., the mobile P2P network among mobile users and the transmission network between users and servers. In SALS, every mobile user periodically shares his unique CZ with others via the P2P communication. Based on the received CZs from neighbors and the knowledge of semantic locations, each user can then ameliorate his CZ for better

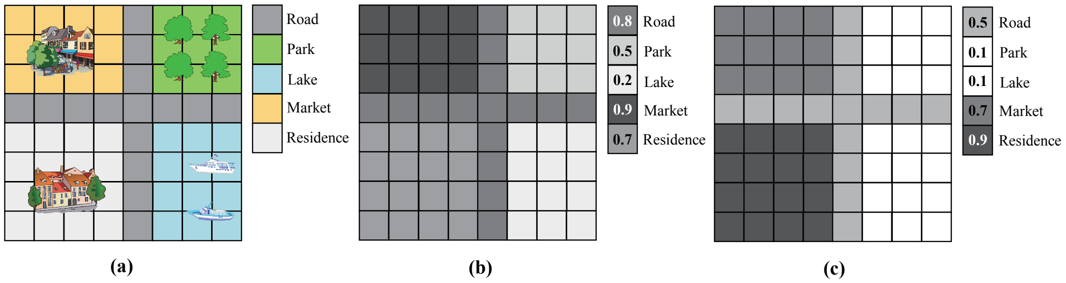


Figure 2: The representation of semantic location based on grid map: (a) semantic location feature identification; (b) semantic location coefficient values at daytime; (c) semantic location coefficient values at night

privacy protection. When accessing location-based social applications, these CZs will be used instead of users' actual locations.

Mobile users carry mobile devices, e.g., smartphones and tablets, which have positioning functionality, e.g., GPS, to allow them accurately determine their geographical positions. Mobile users can access the Internet via the Wi-Fi access points or base stations. The **Mobile P2P network** is a highly ad-hoc network which allows mobile users to communicate with each other via wireless transmission protocols, e.g., Bluetooth or IEEE 802.11, or ad hoc network routing protocols, e.g., LANMAR [11]. This network is self-organized and does not rely on any servers or fixed communication infrastructure. In SALS, mobile users can share their CZs with each other by utilizing this network.

MSN Servers are providing various location-based social applications. These servers manage both the users' personal profiles and their social relationships. In this paper, we assume that these servers are equipped with privacy-preserving query processing engines [4] which can process the location-based queries based on coarse locations. The **Transmission network** is responsible for transmitting messages between the mobile users and MSNs. These messages can be either in plain text or in cipher text which is determined by the service agreements.

5. DESIGN OF SALS

In this section, we firstly describe the semantic location affection and then point out the metrics for evaluating the quality of a CZ.

We transform the real-world map into a discrete grid-based space which contains $m \times n$ grids with equal area. We refer to this discrete grid-based space as $GridMap$ henceforth. Each mobile user is placed into one and only one grid, identified by a row-column pair (x, y). Since users do not trust any entities, we enable them to use a cloaking zone to share with other users or servers, so as to achieve the k-anonymity protection. The definition of CZ is shown below.

Definition 1. The cloaking zone, denotes as CZ, is a rectangle region and can be defined as a 4-tuple record

$$CZ = (id, t, g_{ul}, g_{br})$$

where id indicates the unique ID of a mobile user, t is the generation time of the zone, and g_{ul} and g_{br} represent the upper-left and bottom-right grids of the zone which can be identified by using the coordinates (x_{ul}, y_{ul}) and (x_{br}, y_{br}) in a 2D Euclidean space respectively.

As an example of the semantic location affection, as shown in Figure 2(a), the $GridMap$ has five semantic types, namely road, park, lake, market and residence, which represent the semantic features of the locations in the real world respectively. For each grid in the $GridMap$, we assume that it is small and flexible enough to be assigned to only one semantic feature. For example, the 3×4 grids on the upper-left corner of $GridMap$ are all of semantic feature "Market". In practice, a mobile user can assign each type of semantic feature with a coefficient, denoted as $C(location)$, which can roughly measure the likelihood that other users may appear in that semantic location. As shown in Figure 2(b), we have $C(lake)$ with a relatively small value 0.2 because few people with mobile devices can appear on a lake except for those on boats. We also notice that $C(market) > C(lake)$ since the probability for a user to be located in a market is much greater than on a lake. Moreover, a mobile user can also adjust these coefficients based on some factors, e.g., time and weather. For example, Figure 2(b) shows the coefficient values in daytime whilst Figure 2(c) shows these values at night. So we have $C(park) = 0.5$ at daytime and $C(park) = 0.1$ at night, because the background knowledge shows that the park is closed at night. As aforementioned, the *semantic-location attack* can estimate the victim's real location if he does not consider the semantic location affection when generating the CZ.

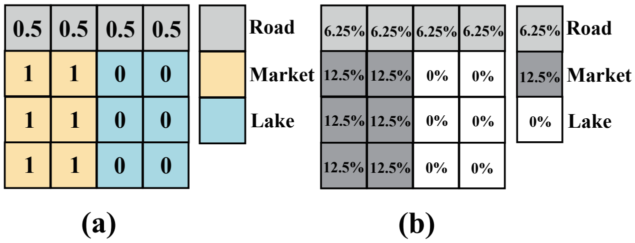
Deducing from the definition of k-anonymity, we conclude that the more anonymous users a CZ covers, the better quality of protection it can provide. Given the semantic location affection, it is wise for a CZ to carefully choose some grids to cover based on the probability that users will appear in that grid. Hence, when user Alice receives a shared CZ from user Bob, Alice can analyze the probability that Bob may be located in each grid inside the CZ. From the other side, it is also the probability that an attacker, when receiving a CZ, can get about this user at a particular grid g. Thus, we define this probability value for each grid as follows.

Definition 2. Given a CZ z, and a grid g inside z, i.e., $g \in z$, the probability for the user who generates z to be located at g, denote as $p(g, z)$, can be calculated by

$$p(g, z) = \frac{C(g)}{\sum_{g' \in z} C(g')}$$

where $C(g)$ is the semantic location coefficient of g.

By the definition, if all the grids covered by a CZ are of the same semantic type, i.e., $C(g)$ are all the same, we can



0.5	0.5	0.5	0.5
1	1	0	0
1	1	0	0
1	1	0	0

6.25%	6.25%	6.25%	6.25%
12.5%	12.5%	0%	0%
12.5%	12.5%	0%	0%
12.5%	12.5%	0%	0%

Figure 3: An example of semantic location affection: (a) the show of semantic location coefficient values; (b) the show of the probability value p(g,z).

conclude that a user's position follows uniform distribution over these grids and the values of $p(g,z)$ all equal to $1/N$ where N is the number of grids inside the CZ. Otherwise, the user's position distribution is nonuniform. For the example in Figure 3, the values of semantic location coefficients are shown in Figure 3(a) from which we know that the lake is unreachable for users but the market is crowded by users. Hence, as shown in Figure 3(b), the probability value $p(g,z)$ for the grids in the lake is 0, whilst the $p(g,z)$ in the market area are as large as 12.5% which means the user has a chance of 12.5% to be located in that grid. Note that the value of $p(g,z)$ ranges in $[0,1)$.

Theorem 1. the sum of $p(g,z)$ is 1, i.e.,

$$\sum_{g\in z} p(g,z) = 1$$

Proof. Suppose that CZ z consists of n grids, i.e., $z = \{g_1, g_2, ..., g_n\}$. Based on Definition 2, we have

$$\begin{aligned}\sum_{g\in z} p(g,z) &= p(g_1,z) + p(g_2,z) + \cdots + p(g_n,z)\\ &= \frac{C(g_1)}{\sum\limits_{g'\in z} C(g')} + \frac{C(g_2)}{\sum\limits_{g'\in z} C(g')} + \cdots + \frac{C(g_n)}{\sum\limits_{g'\in z} C(g')}\\ &= \frac{C(g_1) + C(g_2) + \cdots + C(g_n)}{\sum\limits_{g'\in z} C(g')}\\ &= \frac{\sum\limits_{g\in z} C(g)}{\sum\limits_{g'\in z} C(g')} = 1\end{aligned}$$

□

When a user receives several CZs from his P2P neighbors, these received CZs may overlap with each other. Thus, some grids may be covered by two or more CZs at the same time which will increase the probability that a user may appear in the grid. We refer to this as the total probability value for a grid g, denoted as $P(g)$, which is the sum of probability values contributed by each CZ which covers the gird g.

Definition 3. Let $\mathbb{Z}$ be a set of CZs, i.e., $\mathbb{Z} = \{z_1, z_2, ..., z_k\}$, and let all CZs in $\mathbb{Z}$ cover the grid g, i.e., $g \in \cap_{i=1}^{k} z_i$. The total probability for a user to appear in grid g, denote as $P(g)$, can be calculated as

$$P(g) = \sum_{i=1}^{k} p(g, z_i), \quad where\ g \in z_i$$

Generally speaking, a CZ with a larger area (containing more grids) is more likely to cover more anonymous users thus to achieve a better k-anonymity protection. However, a large CZ also brings several disadvantages. First, a large CZ always indicates low-accuracy location information which will result in low quality of service (QoS). Second, it may cause more communication overhead between users and MSN servers because of the inaccurate locations. Third, to generate a large CZ, a user may need more time and computing resource. Hence, when generating a CZ, it is wise for a CZ to cover some grids with large $P(g)$ value whilst still remaining acceptable size. Therefore, in order to balance the privacy effect against QoS, we defines the *Zone Density* of CZ as a trade-off metrics.

Definition 4. Given a user's own CZ, denoted as z, and a grid g inside the CZ, i.e.,$g \in z$. We measure the quality of z based on the density of anonymous users inside z, denoted as $Den(z)$, which is defined as follows

$$Den(z) = \frac{1 + \sum_{g\in z} P(g)}{N(z)}$$

where $N(z)$ represents the total number of grids inside z which also indicates the area of z.

According to the definition of CZ, $N(z)$ can be calculated by the following formula

$$N(z) = |x_{ul} - x_{br}| \times |y_{ul} - y_{br}|$$

where (x_{ul}, y_{ul}) and (x_{br}, y_{br}) are the coordinates of the upper-left and bottom-right grids of CZ z.

In this paper, we consider both the CZ's area and the covering number of anonymities of the CZ. Hence, we utilize $Den(z)$ as a metric to balance the privacy protection (number of anonymities) against QoS effect (CZ's area). The goal of our algorithm, namely *MaxDen* algorithm, for generating CZ is to find a CZ with maximal possible value of $Den(z)$.

6. IMPLEMENTATION OF SALS

This section firstly gives the overview of SALS, and then shows *MaxDen* algorithm in detail.

6.1 Overview of the Algorithm

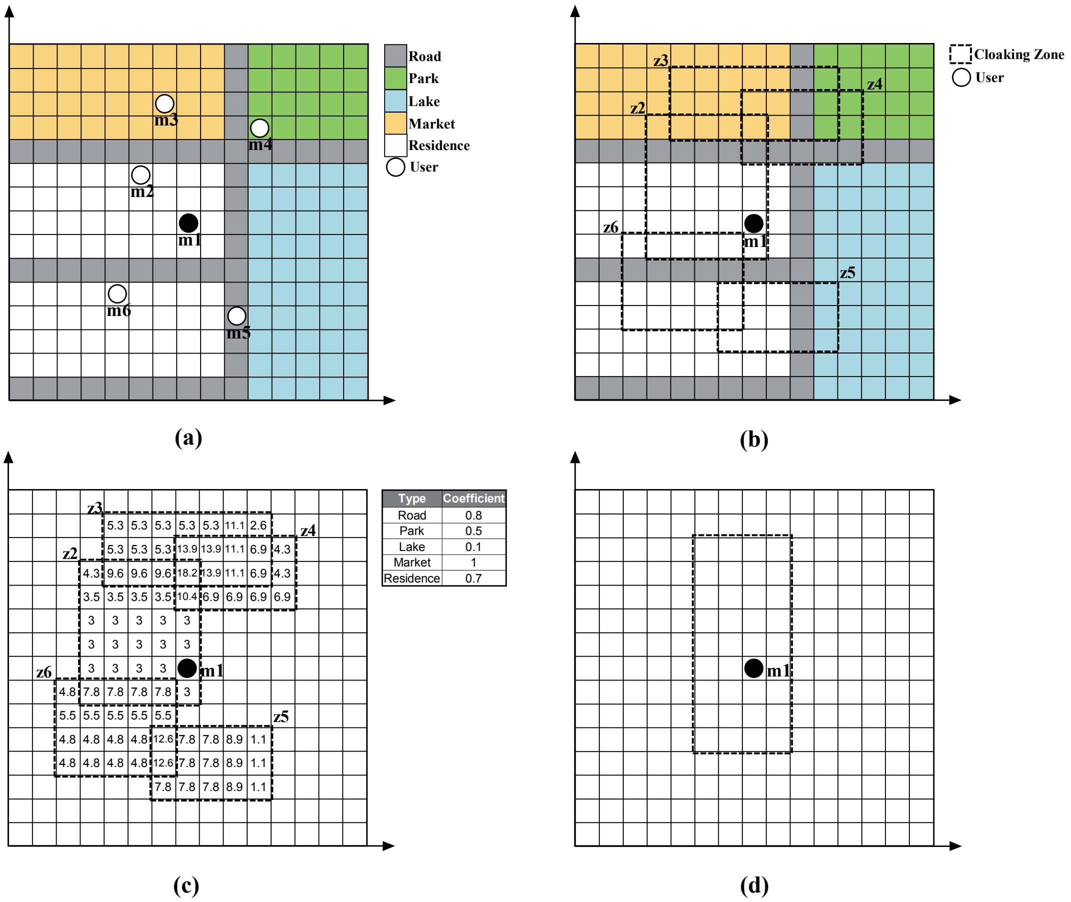
The basic process for generating a user's CZ can be carried out by three phases. Figure 4 shows an example of the process. From Figure 4(a), we have six mobile users $m_1, m_2, ...,$ and m_6 in a 15×15 *GridMap* which contains five types of semantic locations. The following steps show how user m_1 generates his CZ.

Firstly, via the P2P communication technique, user m_1 will receive other five neighbors' shared location information, i.e., five different CZs as shown in Figure 4(b). This phase can be run in several modes, i.e., *on-demand*, *proactive*, or *dual-active* modes [3, 5].

Next, based on these received CZs, user m_1 calculates the $P(g)$ value for each grid around him, as shown in Figure 4(c)[5]. The larger value a grid has, the more likely users may appear in that grid.

Finally, based on these grid values, we then generate m_1's own CZ by using *MaxDen* algorithm. Figure 4(d) gives an example of the generated CZ. The CZ is then shared with other mobile users.

[5] The blank grids have values of 0.



Type	Coefficient
Road	0.8
Park	0.5
Lake	0.1
Market	1
Residence	0.7

(a) (b) (c) (d)

Figure 4: An example to represent the basic process for CZ generating.

6.2 MaxDen Algorithms

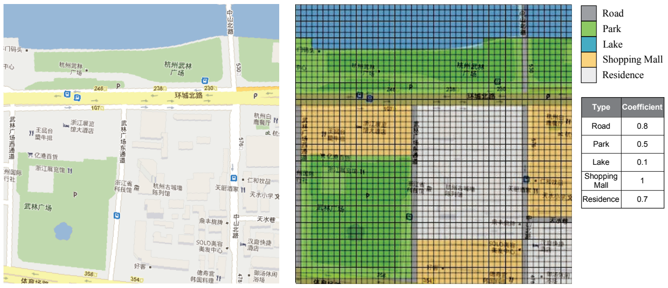
In this part, we describe *MaxDen* algorithm for generating CZ in detail. This algorithm can guarantee to generate a CZ with the largest possible $Den(z)$ value under a user's specified area constraints. In practice, each user may have different requirements on the CZ's area which is defined in user's privacy profile. For example, they can control the area of the CZ by defining the parameters S_{max} and S_{min} which are the maximal and minimal acceptable areas of CZ. The main idea of *MaxDen* algorithm is to enumerate all the possible CZs under the given area limitation so as to select the one with the greatest $Den(z)$. Hence, our algorithm guarantees to output the best result for every scenario. Algorithm 1 shows the pseudo code of *MaxDen* algorithm.

6.3 Security Analysis of SALS

As mentioned in Section 2, an adversary can take either *active* or *passive* attack to violate a user's location privacy. Since most of the previous spatial cloaking solutions cannot defend the *semantic-location attack*, our SALS, however, considers the semantic location affection by assigning each type of semantic locations with different coefficients $C(location)$. By using these coefficients, the distribution of users' locations is no longer uniform but follows the real world's semantic location context. By calculating the probability of users appearing in one grid, we can adjust the CZ's shape and position so as to avoid covering those grids that

Algorithm 1: MaxDen Algorithm

```
Input: $\mathbb{Z}$: a set of CZs, ($S_{min}$, $S_{max}$)
Output: $z$
Define:
$g_{ul}$ = upper left grid of $z$;
$g_{br}$ = bottom right grid of $z$;
$g_u$ = the grid which covers the user;
$maxDen$ = record the max $Den(z)$;
Calculate $P(g)$ for each grid based on $\mathbb{Z}$;
foreach possible zone $z'$ in the GridMap do
    if $z'$ does not contain the $g_u$ then
        continue;
    else if $N(z') < S_{min}$ or $N(z') > S_{max}$ then
        continue;
    calculate the $Den(z')$;
    if $Den(z') > maxDen$ then
        $g_{ul}$ = upper left grid of $z'$;
        $g_{br}$ = bottom right grid of $z'$;
        $maxDen$ = $Den(z')$;
    end
end
return $z = (id, systemtime, g_{ul}, g_{br})$;
```



Type	Coefficient
Road	0.8
Park	0.5
Lake	0.1
Shopping Mall	1
Residence	0.7

(a) (b)

Figure 5: A part of the map of Hangzhou and its GridMap with semantic location context.

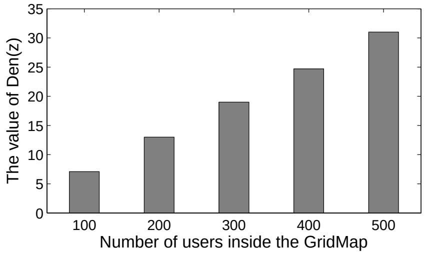


Figure 6: Comparison of Den(z) against user number.

are hard to reach. Since all users follow this generating rule, their CZs are apt to cover some crowded semantic locations after several rounds of generating process. Therefore, it is impossible for an attacker to initiate a *semantic-location attack* on a user's CZ by pruning as many grids as before to narrow down the user's real location range because most of the grids inside the CZ have large and almost similar $C(location)$ values.

The key point for a successful *social-relation-disclosure attack* lies on the social-relation knowledge that an attacker has gathered. Based on these knowledge, an attacker can reveal a victim's real location by relating to his friends' locations. In an MSN environment, since these knowledge are public online, an attacker cannot be prevented from gathering a victim's social relationships. However, our SALS can defend this attack to a certain degree even if the attacker has known the location information of the victim's friends. This is because the CZs shared by everyone are blurred so that the attacker can mine few knowledge from the CZs but only some vague information. The *center-of-zone attack* and *border-of-zone attack* are effective for those algorithms of which the generated CZs often contain the users at the center or boundary of the CZs. Because the goal of *MaxDen* algorithm is to achieve the largest possible $Den(z)$ value, it will not directly affect a user's position in the CZ. As aforementioned, SALS is not designed to defend the *active attacks*, e.g., *region-probing attack* and *location-flooding attack*. In the future work, we can introduce some message-filtering strategies to reject these malicious messages.

7. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the *MaxDen* algorithm from the following three aspects, (1) average $Den(z)$ value, (2) average area of CZ, and (3) the defending efficiency against *semantic-location attack*. For the experimental setup, we choose a part of the map of Hangzhou City in China as the semantic location background, as shown in Figure 5(a). We build up a 50×50 $GridMap$ in which each grid is a square with an area of 10×10 m^2. The semantic context of each grid and the semantic location coefficients are shown in Figure 5(b). The number of simulated mobile users inside the $GridMap$ increases from 100 to 500 for different scenarios. The area constraint for each CZ is $[64, 400]$ grids. The testbed is implemented by Java and run on a PC with Intel Core 2 Quad CPU at 2.4GHz and 1.96GB RAM.

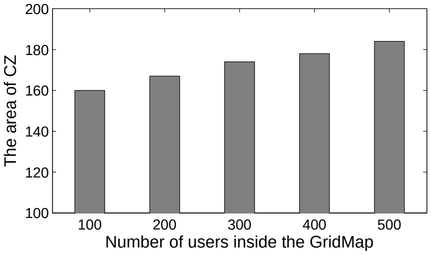


Figure 7: Comparison of CZ's area against user number.

Table 1: The performance against semantic-location attack

Semantic Type	Covering probability	Area ratio
Road	78.2%	11.3%
Park	36.6%	14.2%
Lake	2.6%	0.4%
Shopping Mall	88.4%	45.4%
Residence	48.8%	28.7%

Figure 6 shows that the value of $Den(z)$ grows steadily with the increase of user number in the given $GridMap$, i.e., from 100 to 500. This is because the overall user density of the $GridMap$ is increasing, so is the $Den(z)$ value of the CZ. From Figure 7 we can also observe that the size of CZ is growing with the user number. This is because with more users in the $GridMap$, some small regions in the $GridMap$ will have high user density. Therefore, the CZs whose areas are relatively small are tend to expand their areas to cover these high density regions. As a result, the average size of CZs will increase steadily.

For defending the *semantic-location attack*, the key is to stop attackers from pruning the CZ's area which also means that there should not exist many grids in the CZ whose $C(location)$ values are relatively small. Thus, a well generated CZ is the one in which most of the grids' $C(location)$ values are about the same. In this experiment, we run 500 times of simulation, and calculate the average probability of each type of semantic locations being covered by the generated CZ, and the mean ratio of the area of each type of semantic locations versus the CZ area. In Table 1, the results show that 88.4% of the CZs cover at least one grid with the semantic type of *Shopping Mall*, whereas only 2.6% of the CZs cover the semantic type of *Lake*. Moreover, the area of girds of *Shopping Mall* takes up as many as 45.4% of the total area of CZ, however, the *Lake* only takes up 0.4% of the total area. Therefore, the result verifies that our algorithm can effectively prevent the *semantic-location attack*.

8. CONCLUSION

In this paper, we have proposed SALS, a semantics-aware location sharing framework based on cloaking zone, for preserving mobile users' location privacy in a mobile social network environment. In summary, we have made the following contributions. (1) We have considered the unique characteristic of an MSN environment without assuming any trustworthy entities. As a result, we have enabled users to share CZs rather than their exact locations to any entities. (2) We have taken into account the semantic locations affection when generating a CZ. (3) We have summarized a set of privacy attacks and also evaluated our algorithm against these attacks. The results have shown that our algorithm can well defend the *semantic-location attack*. In respect to the future work, we plan to conduct our experiments in a real application environment. We also plan to investigate the solutions against active attacks.

9. ACKNOWLEDGMENTS

This research is partly supported by the MOE-Intel IT Research Fund of China under Grant No. MOE-INTEL-11-06.

10. REFERENCES

[1] A. Beresford and F. Stajano. Location privacy in pervasive computing. *Pervasive Computing, IEEE*, 2(1):46–55, 2003.

[2] W. Chang, J. Wu, and C. C. Tan. Enhancing mobile social network privacy. In *Proceedings of the IEEE Global Communications Conference (Globecom)*, pages 1 –5, 2011.

[3] Y. Che, Q. Yang, and X. Hong. A dual-active spatial cloaking algorithm for location privacy preserving in mobile peer-to-peer networks. In *Proceedings of the 2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 2543–2548, April 2012.

[4] C. Y. Chow, M. F. Mokbel, and W. G. Aref. Casper*: Query processing for location services without compromising privacy. *ACM Transactions on Database Systems (TODS)*, 34(4):24:1–24:48, 2009.

[5] C.-Y. Chow, M. F. Mokbel, and X. Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based service. In *Proceedings of 14th ACM International Symposium on Geographic Information Systems (ACM-GIS)*, pages 171–178, 2006.

[6] C.-Y. Chow, M. F. Mokbel, and X. Liu. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments. *Geoinformatica*, 15(2):351–380, April 2011.

[7] M. Damiani, E. Bertino, and C. Silvestri. The probe framework for the personalized cloaking of private locations. *Transactions on Data Privacy*, 3(2):123–148, 2010.

[8] M. L. Damiani, C. Silvestri, and E. Bertino. Fine-grained cloaking of sensitive positions in location-sharing applications. *IEEE Pervasive Computing*, 10(4):64 –72, 2011.

[9] M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the First International Conference on Mobile Systems, Applications and Services (MobiSys)*, pages 31–42, 2003.

[10] M. F. Mokbel, C. Y. Chow, and W. G. Aref. The new Casper: Query processing for location services without compromising privacy. In *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB)*, pages 763–774, 2006.

[11] G. Pei, M. Gerla, and X. Hong. LANMAR: Landmark routing for large scale wireless ad hoc networks with group mobility. In *Proceedings of 1st ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 11–18, 2000.

[12] K. P. N. Puttaswamy and B. Y. Zhao. Preserving privacy in location-based mobile social applications. In *Proceedings of the 11th International Workshop on Mobile Computing Systems & Applications (HotMobile)*, pages 1–6, 2010.

[13] W. Wei, F. Xu, and Q. Li. MobiShare: Flexible privacy-preserving location sharing in mobile online social networks. In *Proceedings of the 31st IEEE International Conference on Computer Communications (INFOCOM)*, pages 2616–2620, 2012.

[14] X. Wu, J. Liu, X. Hong, and E. Bertino. Anonymous geo-forwarding in manets through location cloaking. *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 19(10):1297–1309, 2008.

[15] G. Zhong, I. Goldberg, and U. Hengartner. Louis, lester and pierre: three protocols for location privacy. In *Proceedings of the 7th Workshop on Privacy Enhancing Technologies (PET)*, pages 62–76, 2007.