An Anonymous Communication Technique using Dummies for Location-based Services

Hidetoshi Kido[†] Yutaka Yanagisawa[††] Tetsuji Satoh[†,††]
†*Graduate School of Information Science and Technology, Osaka University*
††*NTT Communication Science Laboratories, NTT Corporation*
h-kido@ist.osaka-u.ac.jp yutaka@cslab.kecl.ntt.co.jp satoh.tetsuji@lab.ntt.co.jp

Abstract

Recently, highly accurate positioning devices enable us to provide various types of location-based services. On the other hand, because such position data include deeply personal information, the protection of location privacy is one of the most significant problems in location-based services. In this paper, we propose an anonymous communication technique to protect the location privacy of the users of location-based services. In our proposed technique, such users generate several false position data (dummies) to send to service providers with the true position data of users. Because service providers cannot distinguish the true position data, user location privacy is protected. We also describe a cost reduction technique for communication between a client and a server. Moreover, we conducted performance study experiments on our proposed technique using practical position data. As a result of the experiments, we observed that our proposed technique protects the location privacy of people and can sufficiently reduce communication costs so that our communication techniques can be applied in practical location-based services.

1. Introduction

In recent years, based on sensing technology developments, we can use highly accurate positioning devices such as GPS [6] to obtain the position data of moving objects. Such position data is used in various types of location-based services (LBS) [12]. For example, LBSs provide the nearest restaurant information to users, including location, menu, hours of operation, and so on.

In LBSs, a user generally must send true position data to a service provider, who stores the data in a database. After sending the data, a user cannot delete or modify it. In other words, users cannot prevent service providers from analyzing motion patterns using the stored true position data [3]. To avoid this problem, it is necessary to develop a system to prevent the service provider from learning the user's true position data.

We propose a new anonymous communication technique to protect the location privacy of people using LBSs. In our proposed technique, a user sends true position data with several false position data ('dummies') to a service provider, who creates a reply message for each received position data. The user simply extracts the necessary information from the reply message. In this manner, even if the service provider stores the set of position data, it cannot distinguish the true position data from the set of position data. To apply our anonymous communication technique in LBSs, we discuss the following two important issues:

- Realistic dummy movements
- Reduction of communication costs

Moreover, we explain our proposed anonymous communication technique by defining four evaluation functions based on *Anonymity Set* [9] that evaluates the anonymity of positions. To evaluate our technique, we implemented a simulation system experiment using 39 rickshaw trajectories in Nara City, Japan. To evaluate cost reduction techniques for communication, we also did another experiment using the GeoLink Kyoto [1] service whose results showed that our proposed techniques protect location privacy. From the results of the experiments, we conclude that our technique can be applied into practical LBSs.

The rest of our paper is organized as follows. Section 2 describes location privacy and anonymity. Section 3 describes an *Anonymity Set* and defines evaluation function $AS(i)$. Our proposed anonymous communication and cost reduction techniques are presented in Sections 4 and 5. In Sections 6 and 7, we describe some performance studies of our proposed techniques. Finally, we review related work and offer some conclusions.

2. Location privacy

Beresford and Stajano defined location privacy as "the ability to prevent other parties from learning one's current or past location" [3]. They also said that a system that can

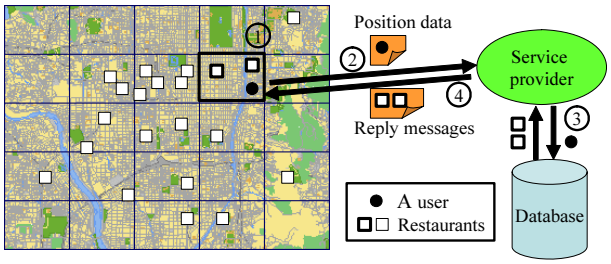


Figure 1. Example of LBS.

obtain position data invades location privacy. In this section, we describe location-based services that protect location privacy. After that we define location anonymity.

2.1. Location privacy for LBSs

We discuss privacy protection for LBSs. LBSs exploit knowledge about where users are located. Figure 1 shows a general example of an LBS. The service's procedure from beginning to end is as follows:

1. An LBS user obtains the true position data of a user using a positioning device such as GPS.
2. The user sends the position data to a service provider.
3. The service provider creates a reply message that responds to the received position data and sends it to the user.
4. The user receives a reply message.

In such a service, the sent message of the user comprises at least a user ID and the true position data. In this paper, we assume that a user ID cannot be connected to privacy information because of pseudonyms. However, even if a user ID is hidden, privacy may be invaded by the position data. Here we show an example.

An LBS gives a user information about when buses will arrive at the nearest stop in a particular vicinity. For example, a person goes to a clinic every week and uses this service at his house and the clinic each time. If such position data are accumulated and analyzed, a staff member or a patient of the clinic may learn the person's address.

This example illustrates that based on position data, location privacy can be invaded. To protect it, service providers must be prevented from learning the true position of users. In other words, it is necessary to anonymize the position data. In this paper, we call the masking of the position data "location anonymity".

2.2. Definition of location anonymity

Pfitzmann and Kohntopp defined "anonymity" as "the state of being not identifiable within a set of subjects."[9] There are many researches about generic anonymous communication, such as Crowds [10] and Onion Routing [7], which anonymize a person by asking: "who sent this data?" However, to date position data anonymity in an LBS has had few discussions that anonymize a location, such as: "where did this data come from?"

We discuss the definition of location anonymity by considering two requirements to enhance location anonymity in an LBS. All LBS users can successfully anonymize their identity:

- **Ubiquity**
  Ubiquity means that subjects exist in an entire area. When all users live in the same region, the service provider can specify users. On the other hand, when users live in various regions, the service provider has difficulty specifying users. Thus, Ubiquity enhances the location anonymity of users in an entire area.

- **Congestion**
  Congestion means that a large number of subjects exists in a region, an idea originated from *k-anonymous* proposed by Gruteser and Grunwald [8]. Users send position data to service provider in a region. When a large number of users live in the region, the service provider has difficulty specifying them. Thus, Congestion enhances their location anonymity in the region.

Ubiquity guarantees the location anonymity of every user. On the other hand, Congestion guarantees location anonymity of local users in a specified cell. Thus, we believe that Ubiquity is more significant than Congestion.

However, there is a case that even if Ubiquity is high, location anonymity is low. When users stay as in Figure 3 (c), a user whom a observer wants to detect often stays in a crowded region. Thus, we consider another requirement: Uniformity.

- **Uniformity**
  Uniformity means that each distributed region includes the same number of users. When an LBS satisfies Uniformity, no users have low location anonymity. In other words, LBS users who satisfy Uniformity have higher location anonymity than LBS users without Uniformity.

In Section 4, we describe a new technique to provide LBSs with Ubiquity, Congestion, and Uniformity.

3. Enhanced Anonymity Set for LBS

Anonymity Set, a measure that evaluates anonymity, was originated by Chaum [5]. Pfitzmann and Kohntopp define an *Anonymity Set* as "the set of all possible subjects."[9] In this section, we produce new evaluation functions to quantify location anonymity based on *Anonymity Set*.

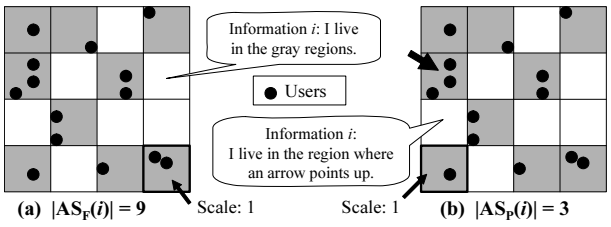


Figure 2. Examples of $AS(i)$.

3.1. Enhanced Anonymity Set

The idea of *Anonymity Set* can be applied to the field of generic anonymous communication. We extend it to location anonimization techniques in LBSs. The extended definition of an original *Anonymity Set* is the set of all subjects determined by information about position. Here, we symbolically define the enhanced *Anonymity Set* to propose evaluation functions.

First, we define the following symbols.

a	:	a subject
A	:	a set of subjects, $A = \{a_1, a_2, \ldots, a_n\}$
i	:	information about A
I	:	a set of information
$\|A\|$	:	cardinality of A
$\hat{A}$	:	power set of A (2^A)

Each i is represented as a *sentence* that shows information that limits a set belonging to A. For example, assume that A is a set of people. When it provides i to each element included in A who live in Japan, i restricts the set to all people living on earth to a set of all people living in Japan.

Next, based on the symbols, we notate function $AS(i)(i \in I)$ and the cardinality of the number of elements as follows:

$$\begin{aligned} AS(i) &= 2^A = \hat{A} \quad (AS : I \rightarrow \hat{A}) \\ |AS(i)| &= |\hat{A}|. \end{aligned}$$

3.2. Evaluation function based on $AS(i)$

We propose the following two functions to evaluate location anonymity.

- $AS_F(i)$:
  $AS_F(i)$ is a function that returns α_F, which is a set of regions specified by i. $|AS_F(i)|$ denotes the number of α_F and shows the total scale of α_F or the number of α_F if the regions are of the same scale. $AS_F(i)$ can be defined as follows, where each r_j is a region:

$$\begin{aligned} A_F &= \{r_1, r_2, \ldots, r_m\} \subset A \quad (\forall r_j \in A_F) \\ |A_F| &= |\{r_1, r_2, \ldots, r_m\}| = m \\ AS_F(i) &= \alpha_F (\in \hat{A}_F) \quad (AS_F : I \rightarrow \hat{A}_F) \\ |AS_F(i)| &= |\alpha_F|. \end{aligned}$$

Table 1. Example of location anonymity for Figure 3.

		(a)	(b)	(c)	(d)	(e)
Ubiquity	**F**	High	Low	High	Low	High
Congestion	**P**	Low	High	Mid	Low	High
Uniformity	$\mathrm{Var}(\mathbf{P})$	High	Low	Low	High	High

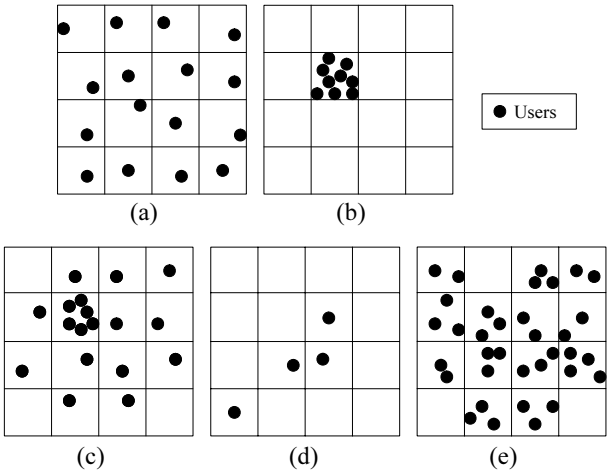


Figure 3. Example of distribution of position data.

Figure 2 (a) shows an example of $|AS_F(i)|$. The scale of the region is one. In Figure 2 (a), $|AS_F(i)|$ is nine when i is provided with "I live in the gray regions."

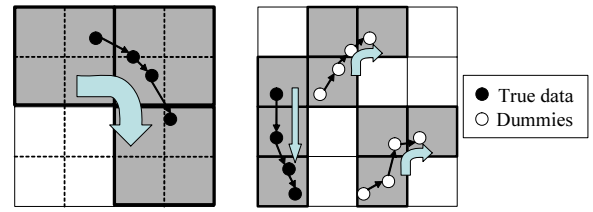
- $AS_P(i)$:
  $AS_P(i)$ is a function that returns α_P, which is a set of persons limited by i. $|AS_P(i)|$ denotes the number of α_P and shows the number of α_P. $AS_P(i)$ can be defined as follows, where each p_j is a person:

$$\begin{aligned} A_P &= \{p_1, p_2, \ldots, p_m\} \subset A \quad (\forall p_j \in A_P) \\ |A_P| &= |\{p_1, p_2, \ldots, p_m\}| = m \\ AS_P(i) &= \alpha_P (\in \hat{A}_P) \quad (AS_P : I \rightarrow \hat{A}_P) \\ |AS_P(i)| &= |\alpha_P|. \end{aligned}$$

  Figure 2 (b) shows an example of $|AS_P(i)|$. In Figure 2 (b), $|AS_P(i)|$ is three when i is provided with "I live in the region where an arrow points up."

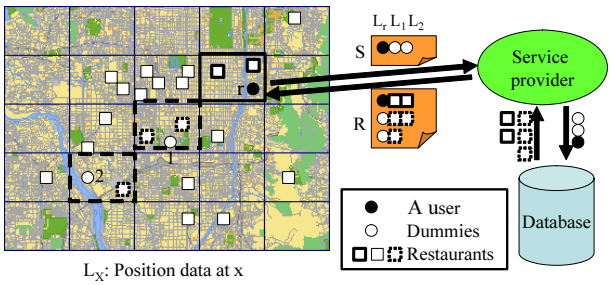
3.3. Quantification of location anonymity

We describe the quantification of location anonymity using enhanced *Anonymity Set* by defining two more symbols: **F** and **P**. **F** denotes $|AS_F(i)|$ in which i is provided with information that determines multiple regions. Thus, **F** means a scale of all regions where users stay. **P** denotes $|AS_P(i)|$ in which i is provided with information that determines a specific region. Thus, **P** means the number of users in a



(a): Accuracy reduction (b): Dummy generation

Figure 4. Two types of anonymous communication techniques for LBS.



L_X: Position data at x

Figure 5. Example of anonymous LBS using our technique.

specific region. Note that all areas that provide the service are divided into regions, as shown in the map of Figure 1. The regions are the same scale as the accuracy of the position data.

Now, we describe the relationship between symbols **F** and **P**, and the three elements for an anonymous LBS: Ubiquity, Congestion, and Uniformity, as shown in 2.2. As explanation, Figure 3 shows five examples of the distribution of position data. Also, Table 1 shows the degree of location anonymity for the examples as three states: High, Mid, or Low.

- Ubiquity—**F**
  F corresponds to Ubiquity. In other words, an increase of **F** enhances location anonymity. As shown in Figure 3 and Table 3, when there are many regions where people live, LBSs have Ubiquity, and user location anonymity is high.

- Congestion—**P**
  P corresponds to Congestion. In other words, an increase of **P** enhances location anonymity. In an exception, regions at $\mathbf{P} = 0$ are not considered because no people live in that region. As shown in Figure 3 and Table 3, the region where many people live has Congestion, and user location anonymity in the region is high.

Uniformity can be defined using the variance of **P**.

- Uniformity—the variance of **P** (notated $\mathrm{Var}(\mathbf{P})$)
  If $\mathrm{Var}(\mathbf{P})$ is low, a variation of **P** in each region is low, too. Thus, if $\mathrm{Var}(\mathbf{P})$ is low, then a LBS with Ubiquity has also Uniformity. As shown in Figure 3 and Table 3, when each region includes the same number of users, an LBS has Uniformity. However, note that if an LBS does not have Ubiquity, as shown in Figure 3 (d), it does not enhance user location anonymity.

4. Anonymous communication technique

We propose a new anonymous communication technique for LBS.

Gruteser and Grunwald proposed an anonymous usage of a location-based service [8]. In this usage, a user does not send his position data without modification obtained by GPS to the service provider but sends it with accurate information, which is reduced. Figure 4 (a) shows an example of such usage. In Figure 4, the user does not send, "I live at a point," but instead sends "I live in a gray region." The service provider can only learn vague details of the position of the user in the usage. Thus, the usage enhances location anonymity.

However, the approach has the following problem: observers can easily comprehend user moves when tracing data for several minutes because the chain of the position data creates a rough trajectory, as in Figure 4 (a). Moreover, if the accuracy of position data is reduced, the accuracy of service will also be reduced. In this section, we describe the basic idea of our proposed anonymous communication technique that tackles the problem in 4.1. Then in 4.2, we propose dummy generation algorithms.

4.1. Basic idea

Gruteser's usage has a problem: an adversary can easily comprehend the moves of users when an observer traces the data for several minutes [8]. To address the problem, there must be some different position data that cannot be distinct from a user's true position data.

Based on that idea, we propose a new anonymous communication technique for LBSs in which a user sends position data including noise to the service provider. The noise consists of a set of false position data called 'dummies.' Here, we describe how to use anonymous LBSs with our technique. Figure 5 shows an example of an anonymous LBS using our technique. $L_x = (X_x, Y_x)$ shows position data at x. A requiring message (**S**) from the user to the service provider is defined below:

$$\mathbf{S} = (u, L_1, L_2, \ldots, L_m),$$

where u shows a user ID and $(L_1, l_2, \ldots, L_m)$ shows a set of position data that includes one true position data and $m-1$ dummies. For example, **S** consists of (u, L_r, L_1, L_2) in Figure 5. On the other hand, a service answer message (**R**) from the service provider to the user is defined below:

$$\mathbf{R} = ((L_1, D_1), (L_2, D_2), \ldots, (L_m, D_m)),$$

where $(D_1, D_2, \ldots, D_m)$ shows the contents of the service related to $(L_1, L_2, \ldots, L_m)$. For example, **R** consists of $((L_r, D_r), (L_1, D_1), (L_2, D_2))$ in Figure 5, the service's procedure from the beginning to the end is as follows:

1. The positioning device obtains position data *r* of the user.
2. Dummies are generated at positions *1* and *2*.
3. The user creates a service requiring message **S** that includes position data at *r*, *1*, and *2*; moreover, the user sends **S** to the service provider.
4. The service provider creates service answer message **R** that responds to receiving all position data and sends **R** to the user.
5. The user receives **R** and only selects necessary data from **R**.

The user knows the true position data, but not the service provider. So the service provider cannot distinguish true position data from a set of received position data. In this way, anonymous service is complete.

Figure 4 (b) shows an example of our technique. In Figure 4 (b), the user sends position data at the gray region that is identical to (a), but the user generates two dummies different from (a). The dummies can move in different directions from the true position data. Consequently, comprehending user moves is more difficult. Actually, because the other users simultaneously send position data, the user is more secure.

4.2. Dummy generation algorithm

We describe dummy generating. Dummies must be generated so that they cannot be distinguished from true position data. In some LBSs, as road navigation services, the user must send continuously position data. Generally speaking, the distance each subject can move in a fixed time is limited. If dummies are generated randomly, observers can easily find differences between true position data and dummies. In this case, location anonymity is reduced. To avoid this, the dummy must not behave completely different from the true position data. We present the following two dummy generation algorithms to prevent service providers from finding the true position data. In these algorithms, the locations of the first dummies are decided randomly because the algorithms use the previous location of dummies.

- Moving in a Neighborhood (MN)
  In this algorithm, the next position of the dummy is decided in a neighborhood of the current position of the dummy. Table 2 shows an outline of the algorithm. An example of dummies adapted to the algorithm is illustrated in Figure 6 (a). In this algorithm, the communication device of the user memorizes the previous position of each dummy. Then the device generates dummies around the memory.

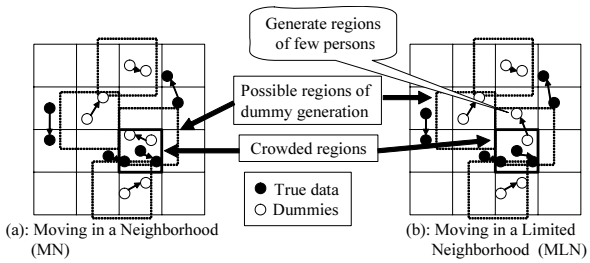


Figure 6. Illustration of two dummy generation algorithms.

- Moving in a Limited Neighborhood (MLN)
  In this algorithm, the next position of the dummy is also decided in the neighborhood of the current position of the dummy. However, the next position is limited by the density of the region. This algorithm is adaptable in cases where the communication device of the user can get the position data of other users. Table 3 shows an outline of the algorithm.
  An example of dummies adapted to the algorithm is illustrated in Figure 6 (b). First, the user device gets the other user's position data. Next, the device generates dummies around the memory that are the same as the MN algorithm. If there are many users in the generated region, the device generates the dummy again. The process is repeated several times.

We define $\mathrm{Shift}(\mathbf{P})$ as a measure that evaluates the two algorithms. $\mathrm{Shift}(\mathbf{P})$ expresses a difference of **P** in each region between times `t` and `t+1`. If the number of persons changes greatly in a region, there is a high possibility that the dummy moves strangely compared with the true position data, creating a risk that observers may find true position data. To that end, it cannot enhance location anonymity. In other words, when $\mathrm{Shift}(\mathbf{P})$ decreases, location anonymity is high.

5. Cost reduction technique

In our anonymous communication technique for LBSs, if the number of dummies increases, location anonymity is enhanced, but communication costs increase. To avoid such increases, we propose a cost reduction technique for our communication technique. In this section, first, we outline communication costs in 5.1. Next we propose a cost reduction technique for requiring messages from a client to a server and answer messages from the server to the client in 5.2 and 5.3.

5.1. Communication cost

Beresford conjectured that the cost of an anonymous technique using dummy users might be too high in real-world services [3]. In our communication techniques, communication cost increases are generated as a side effect to

Table 2. Moving in a Neighborhood (MN) algorithm.

```
// Input:  positions of dummies at t-1
// Output:  positions of dummies at t
// random(x,y):  generate a random number between x and y

struct dummy {
       double x;         // x coordinate
       double y;         // y coordinate
       double t;         // time
};
void MN (double m, int n) {
      struct dummy prev[100], next[100];

      (Assignment prev[] to the Input);
      for (i=1;i<n;i++) {
            next[i]->x = random( (prev[i]->x)-m, (prev[i]->x)+m);
            next[i]->y = random( (prev[i]->y)-m, (prev[i]->y)+m);
            next[i]->t = (prev[i]->t)++;
      }
      (Output the contents of the next[]);
}
```

Table 3. Moving in a Limited Neighborhood (MLN) algorithm.

```
// Input:  positions of dummies at t-1
// Output:  positions of dummies at t
// random(x,y):  generate a random number between x and y
// position(x,y):  return the amount of position data where (x,y,t-1) belongs

struct dummy { (defined in Table 2) };
void MLN (int aveP, double m, int n) {
      struct dummy prev[100], next[100];
      int k = 0; // repeat count (default:3)

      (Assignment prev[] to the Input);
      for (i=1;i<n;i++) {
            next[i]->x = random( (prev[i]->x)-m, (prev[i]->x)+m);
            next[i]->y = random( (prev[i]->y)-m, (prev[i]->y)+m);
            next[i]->t = (prev[i]->t)++;
            if (position(next[i]->x, next[y]->y) > aveP) {
                  if (k<=3) { k++; continue; } else { k=0; }
            }
      }
      (Output the contents of the next[]);
}
```

enhance location anonymity. This is one serious obstacle for practical use. On the other hand, the service provider must create a reply message not only for the true position data but also for the dummy. Since the processing cost for the dummy is quite low, we do not consider processing costs in this paper.

5.2. Requiring messages

A requiring message **S** is shown in 3.3. When dummies are generated, increases in the amount of position data directly cause an increase in the cost of sending messages. In the old way, **S** is constructed as:

$$(u, (X_r, Y_r), (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)),$$

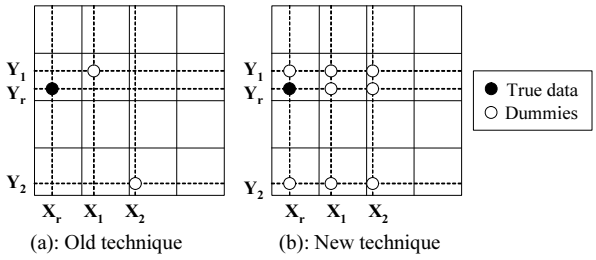


Figure 7. New technique to reduce costs for requiring messages.

where $L_r = (X_r, Y_r)$ is the true position data and $(X_i, Y_i)(i = 1, 2, 3, \ldots n)$ are dummies. In this case, dummies are generated, as in Figure 7 (a). If *n* dummies are

generated, the cost for requiring messages is $O(n)$. In other words, when the number of dummies increases, the cost for requiring messages increases proportionally.

To decrease costs, we propose a new technique **S** constructed as follows:

$$(u, (X_r, X_1, X_2, \ldots, X_n), (Y_r, Y_1, Y_2, \ldots, Y_n)).$$

In the technique, position data are divided into sets of coordinate X and Y that include true position data X_r and Y_r. The service provider recognizes all of their combinations as position data illustrated in Figure 7. In other words, a set of (n + n) data consists of n^2 position data. The communication cost for requiring messages with the technique is $O(\log n)$, with more decreases than previous methods.

5.3. Answer messages

The basic idea of cost reduction for an answer message is a reduction of the number of data using keywords.

An answer message **R** is shown in 3.3. In **R**, service content D_j is comprised of the name and attributes related to a kind of service. D_j is generally constructed as follows:

$$(name_1, URL_1, address_1), (name_2, URL_2, address_2) \cdots.$$

Each position has a different number of tuples $(name_j, URL_j, address_j)$. If there are too many tuples in an answer message, the service provider can request that the user send more information to limit the amount of answer messages.

Here, we consider reducing the costs for answer messages. Position data costs cannot be decreased because all position data connect to the service data. To remove unnecessary information in the service provider is also impossible because it becomes a key for distinguishing dummies. Therefore, we plan to limit service data using information unrelated to position data. We propose the following four techniques to reduce costs.

- Range limitation:
  If the accuracy of position data increases, the amount of service data connected to the position data decreases. In a restaurant search service, the amount of restaurant information sent to the user is reduced. But in this technique, there is a problem: the degree of location anonymity is reduced at the same time.

- Category limitation:
  Most search services are hierarchized by categories that limit kinds or names in the services. The user can send information to limit categories if the cost of answer messages is too high. Limiting data by category can reduce the amount of service data. This technique greatly affects services with many categories.

- Setting keywords:
  Most LBSs include information by text. These services can limit answer message cost with keywords sent by the user if the cost is too high.

- Removal of unnecessary data:
  In LBSs, users often receive unnecessary data. We show an example of a restaurant search service as an explanation. Suppose a person wants to go to a popular Japanese restaurant he found in a magazine. He uses this service to learn its precise location. In this situation, menu information of the restaurant is unnecessary. In this way, the technique limits the data the user considers unnecessary.

In these techniques, the user needs to send information unrelated to position data to the service provider. The user can add service limitation information to the requiring message. The cost of the information is too small to be ignored.

6. Evaluation of an anonymous communication technique

To evaluate our technique shown in Section 4, we implemented a simulation system. Using the system we experimented with actual trajectory data. In this section, we show experimental outlines and results.

6.1. Settings

For simplification, we added the following two assumptions:

- All users generated the same number of dummies. That is, if a user generates two dummies, other users also generate two.

- Except for position data, the user does not send personal information.

First, to examine how dummy generation effectively enhances location anonymity, we did the following experiment:

- Comparison of location anonymity and number of dummies.

It is important to learn the effects of location anonymity when the number of dummies increases. Ubiquity **F** is a most suitable measure to evaluate location anonymity.

Next, to evaluate our proposed dummy generation algorithm, we did two more experiments:

- Relationship between dummy generation algorithms and $\mathrm{Shift}(\mathbf{P})$.

- Relationship between dummy generation algorithms and $\mathrm{Var}(\mathbf{P})$.

These experiments showed that our proposed dummy generation algorithm is better than the random generation of dummies.

We implemented a simulation system for the experiments that can deal with and display coordinates x and y

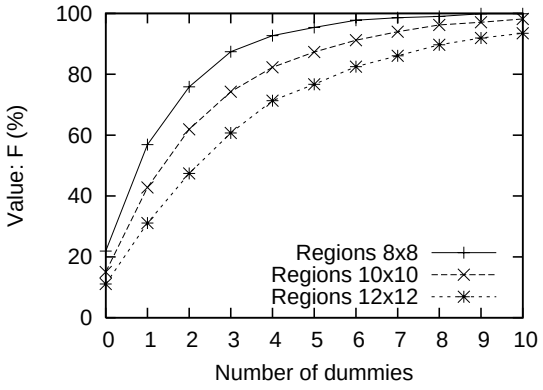


Figure 8. Comparison of location anonymity and number of dummies.

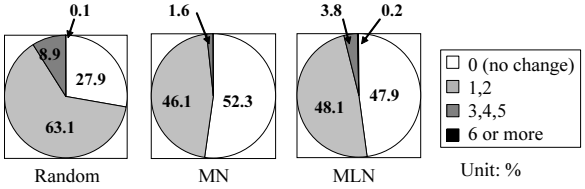


Figure 9. Relationship between dummy generation algorithms and $\mathrm{Shift}(\mathbf{P})$.

and time t. Moreover, the system has a module that generates dummies based on true position data. We can calculate the values of **F**, **P**, $\mathrm{Var}(\mathbf{P})$, and $\mathrm{Shift}(\mathbf{P})$ using the system. For the experiments, we gave the system 39 rickshaw trajectories from Nara City, Japan.

6.2. Results

Figure 8 shows that when the number of dummies increases, location anonymity also increases. We set the number of dummies between 0 and 10, and a unit of **F** was the percentage of the entire area whose scale is about 10×10 *km*, and the number of regions was 8×8, 10×10, or 12×12. Figure 8 shows that for location anonymity, a setting in which one dummy is generated in 8×8 regions is higher than another setting in which a dummy is not generated in 12×12 regions. In other words, the dummy generation technique enhances location anonymity more effectively than the accuracy reduction technique. Moreover, as expected, the more dummies, the larger the value of **F**. As shown in Figure 8, if a user achieves 80% of **F**, we conclude that the user needs three dummies in 8×8 regions, four dummies in 10×10 regions, and six dummies in 12×12 regions.

Figure 9 shows the relationship between dummy generation algorithms and $\mathrm{Shift}(\mathbf{P})$. This figure shows not only the two proposed algorithms but also a random algorithm that randomly generates dummies because of comparisons. We set the number of regions at 10×10 and the number of dummies at three. The results reveal that both algorithms have less $\mathrm{Shift}(\mathbf{P})$ than the random algorithm. Thus, we conclude that the two proposed algorithms are more effective than the random. On the other hand, when the two algorithms are compared, the MN algorithm is slightly better because the MLN algorithm increases opportunities that the dummy crosses between two regions.

Table 4. Relationship between dummy generation algorithms and $\mathrm{Var}(\mathbf{P})$.

	Random	MN	MLN
$\mathrm{Var}(\mathbf{P})$	2.43	2.26	2.08

Table 4 shows the relationship between the dummy generation algorithms and $\mathrm{Var}(\mathbf{P})$. The situation of this experiment is the same as Figure 9. In Table 4, it turns out that both algorithms have less $\mathrm{Var}(\mathbf{P})$ than the random algorithm. Therefore, we conclude that the two proposal algorithms are more secure than the random. On the other hand, different from Figure 9, when the two algorithms are compared, the MLN algorithm is slightly better, showing that the MLN algorithm works as expected.

7. Evaluation of cost reduction technique

To evaluate our cost reduction technique shown in Section 5, we conducted another experiment with the GeoLink Kyoto [1] service. In this section, we show experimental outlines and results.

7.1. Settings

To evaluate how they reduce communication costs, we did two more experiments as follows:

- Cost comparisons for requiring messages
- Cost comparisons for answer messages

Each study shows how to reduce costs with our cost reduction techniques. We used the GeoLink Kyoto [1] service (Figure 10) for the experiments. The service presents web pages of various spots in Kyoto and has a database that includes the ID, name, URL, position (latitude, longitude), address, category, and a memo of each spot.

Requiring message **S** includes service limited information in addition to a set of position data. Position data is 32 bit actual numbers per number. The rest of the data of **S** is 16 bytes constant in total. Answer message **R** consists of each sent position data and service data connected to the position data shown in 4.1. The service data consists of all items in the GeoLink service database within the position data. We calculated that the average total cost per spot is 121.964 bytes. So we defined it as 128 bytes including the cost of the position data and the packet header. The position data shows a range of a circle with a two *km* radius. In these

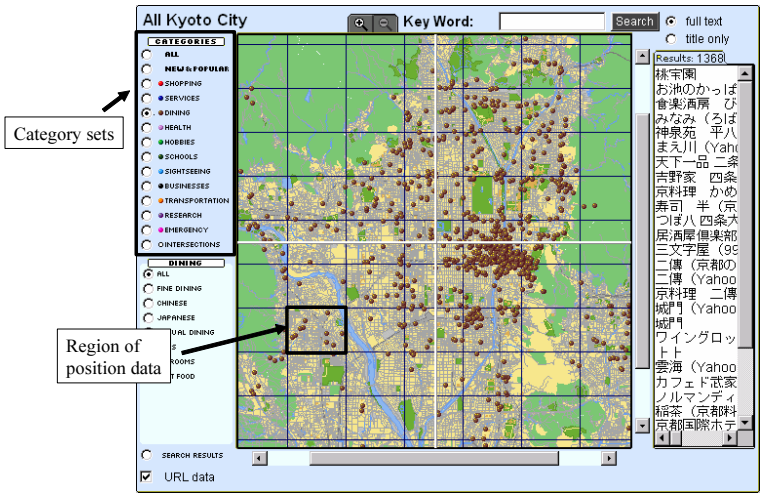


Figure 10. Example of GeoLink Kyoto service (category: dining).

Table 5. Cost comparisons for requiring messages for our techniques.

Number of sending P.D.	Message size [Bytes]	Number of recognizing P.D. without technique	Number of recognizing P.D. with technique
1	24	1	1
2	32	2	4
3	40	3	9
4	48	4	16
5	56	5	25
10	96	10	100
100	816	100	10000

settings, the average number of services per position data is 114.47, and the maximum is 1.067. Therefore, the average answer message cost per position data is as follows:

$$128 \times 114.47 = 14,652.16[\mathrm{Byte}] \simeq 14.7[\mathrm{KByte}].$$

Maximum answer message cost is as follows:

$$128 \times 1,067 = 136,576[\mathrm{Byte}] \simeq 137[\mathrm{KByte}].$$

We used the values in the experiments. The cost shown in the value causes no problem if the number of dummies is low. However, when many users simultaneously generate many dummies, the service will generate a large communication delay.

In answer message experiments, we stored all the data of the GeoLink service in the PostgreSQL [2] database. We made pseudo position data using SQL to query the database. We queried 100 times and calculated averages. We also set categories or keywords using SQL.

7.2. Results

Table 5 shows requiring message cost comparisons between using cost reduction techniques and not. The amount

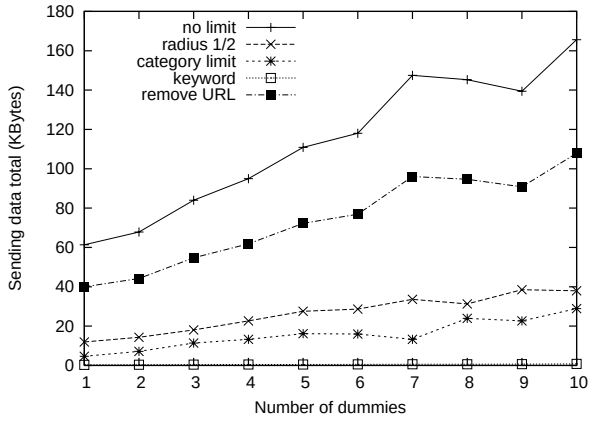


Figure 11. Cost comparisons for answer messages costs for our cost reduction techniques.

of sending position data (P.D.) shows the number of sets of coordinates x and y. The amount of recognizing position data (P.D.) shows the amount of position data received by the service provider. As shown in Table 7, if the amount of sending position data increases by one, message cost increases eight bytes. Our technique forces service providers to recognize position data squares as much as sending position data. We concluded that our communication technique generates dummies at $O(\log n)$ cost when our cost reduction technique is used. Even if a user sends 10,000 position data, the cost of a requiring message is only 816 bytes. Services using our technique are realizable.

Figure 11 shows the average of the total answer message cost for our cost reduction technique. The X axis shows the number of dummies. The range of obtaining data is a circle with a two *km* radius. True position data has a two *km* range in a radius from the Kyoto Prefectural Gym. The leg-

ends of Figure 11 are defined as follows: `no limit` shows not sending limited information, `radius 1/2` shows the radius of the circle to get half of the data, `category limit` shows setting category "dining," `keyword` shows setting keyword "McDonald's," and `remove URL` shows removing URL data. In Figure 11, every technique reduces the cost of answer messages. Especially when setting categories or concrete keywords, the cost greatly decreased because the number of spots is greatly decreased. It is considered that the throughput of wireless communication will become several Mbps in the next several years. We conclude that our techniques can apply to LBSs in the near future.

8. Related work

The field of anonymous communication was originated by Chaum, who described mix networks [4] and the dining cryptographers algorithm [5]. In [4], he proposed an untraceable communication system called the *mix* that used a mail system, digital signatures, and so on. In [5], he also proposed untraceability between sender and recipient and the origin of *Anonymity Set*.

A prior work on location privacy is Mix Zones [3], which is similar to mix networks. In Mix Zones, infrastructure provides an anonymous service using pseudonyms that collects and reorders messages from users within a mix zone to confuse observers. A problem with this system is that there must be enough users in the mix zone to enhance location privacy. Different from our techniques, this system works in buildings or in small blocks in urban areas.

Gruteser and Grunwald proposed another mechanism called spatial and temporal cloaking [8] that conceals a user within a group of *k* people, called *k-anonymous*, which originated from *k-anonymity* [11]. To achieve *k-anonymous*, spatial or temporal accuracy of location information is reduced. However, in this mechanism, when there are too few people in a small area, the accuracy of location information is too low to use for LBSs. In our method, even if users send more accurate location information than this mechanism, location privacy is protected.

On the other hand, there are some anonymous communication works such as Crowds [10] and Onion Routing [7]. Crowds is a re-routing system for anonymous web browsing. A web server cannot identify a request because it resembles a member of the crowd, a set of users who perform the same actions. Onion Routing is an anonymization protocol for an IP network layer. Senders encrypt routing information using the onion routers public key.

9. Conclusions

In this paper, we proposed a new anonymous communication technique for location-based services to protect location privacy using dummies. In the technique, a client system generates several false position data, which the system sends with the true information of the person to the service provider. Even if another person intercepts the data, the person cannot distinguish true position data from an amount of position data. We also proposed a cost reduction technique for communication with our anonymous technique. Moreover, we did experiments with the technique using actual trajectory data and the GeoLink service [1]. Evaluation results using *Anonymity Set* showed that the communication techniques protect location privacy and can be applied in practical LBSs.

As future work, we think that a dummy should draw a trajectory closer to the true position data. Moreover, we need to produce new measures to evaluate location anonymity.

Acknowledgments

Professor Masayuki Murata at the Graduate School of Information Science and Technology, Osaka University, helped our research. We thank him and all the others who assisted us.

References

[1] http://www.digitalcity.gr.jp/openlab/kyoto/map_guide.html.
[2] http://www.postgresql.org/.
[3] A. R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.
[4] D. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 4(2), February 1981.
[5] D. Chaum. The dining cryptographers problem: Unconditional sender and recipient untraceability. *Journal of Cryptology*, 1:65–75, 1988.
[6] I. Getting. The global positioning system. In *IEEE Spectrum*, volume 30, pages 36–47, December 1993.
[7] D. Goldschlag, M. Reed, and P. Syverson. Onion routing for anonymous and private internet connections. *Communications of the ACM (USA)*, 42(2):39–41, 1999.
[8] 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*, pages 31–42, 2003.
[9] A. Pfitzmann and M. Kohntopp. Anonymity, unobservability, and pseudeonymity: a proposal for terminology. In *International workshop on Designing privacy enhancing technologies*, pages 1–9. Springer-Verlag New York, Inc., 2001.
[10] M. Reiter and A. Rubin. Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security*, 1(1):66–92, June 1998.
[11] P. Samarati and L. Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, 1998.
[12] O. Wolfson, P. Sistla, B. Xu, J. Zhou, and S. Chamberlain. DOMINO: Databases fOr MovINg Objects tracking. In *SIGMOD'99 Conference Proceedings*, pages 547–549, 1999.