

# Providing Location Privacy in Assisted Living Environments \*

Yi Ouyang<sup>1</sup>, Yurong Xu<sup>1</sup>, Zhengyi Le<sup>1</sup>, Guanling Chen<sup>2</sup>, Fillia Makedon<sup>1</sup>

<sup>1</sup>Computer Science and Engineering Department  
University of Texas at Arlington, TX  
{yi.ouyang, yurong, zyle, makedon}@UTA.EDU

<sup>2</sup>Computer Science Department  
University of Massachusetts at Lowell, MA  
glchen@cs.uml.edu

## ABSTRACT

While pervasive technology becomes more widely used in assisted living environments, it becomes more important to preserve the privacy of patients being monitored. Location data of patients can be collected through sensors for behavior patterns analysis, and they can also be shared among researchers for further research for early disease diagnosis. However, sharing location information also introduces privacy concerns. A series of consecutive location samples can be considered as a trajectory of a single person, and this may leak private information if obtained by malicious users. In this paper, this problem is discussed and a location randomization algorithm is proposed to protect users' location privacy. We defined privacy metrics according to location privacy and proposed a method using dynamic mix zones to confound trajectories of two or more persons.

## Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

## Keywords

Algorithms, Sensor Networks, Location Privacy

## 1. INTRODUCTION

Wireless devices and pervasive technology becomes more and more widely used in assisted living environments. Wireless sensor networks can be used to monitor movements of patients and collect their medical data simultaneously. Data collected by sensor networks can be transferred back to a storage server or doctor's

---

\*This work was supported in part by the National Science Foundation under award number ITR 0733674, IDM 0733673, and CT-1SG 0716261. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2008 ACM PETRA-7/15/2008 ...\$5.00.

office for further analysis. Realtime monitoring of patients' moving patterns and behavior can help doctors observe the progress of patients' disease and facilitate early disease diagnosis. Location information of the patients in an assisted living environment is very important for behavior monitoring. Location data of a patient can be used to facilitate diagnosing this patient, and can also be used for general medical research when shared by medical researchers. While the behavior monitoring application through pervasive technology is attractive, however, releasing location information to doctors and sharing location information with distributed entities can also introduce privacy concerns. In this paper, we will discuss the privacy issues brought by location information collection in assisted living environments. It is the goal of this work to provide useful location services while preserving the patients' location privacy to the greatest extent possible.

Location information can be collected actively or passively. Active collection often occurs in sensor network tracking applications, in which users' locations are recorded by sensors and sent back to a base station, which is usually also the location server. Passive collection is also common; for example, a patient might transmit her location to a central server. In the active collection case, since locations are collected without the control of users, patients cannot prevent their location information from being collected by location servers and being shared with other applications. In the passive collection case, a patient can decide whether and when to send her location information to the server—thus, the patient can enforce device policies on the location information being sent. However, after location data is received by the server, a patient cannot exert any further control over it. Thus, although enforcing policies on the release of location data can help preserve patients' privacy, methods to preprocess location data in order to preserve patients' privacy are still desirable before these location data are shared with other doctors.

The straightforward way to preserve patients' privacy seems to be to anonymize location data—in other words, to remove any identity information from the location data. However, careful consideration of this idea reveals that even reporting only raw location data without identities is not enough to protect the privacy of patients: because of the continuity of motion data, locations of a single patient can be tracked using various algorithms. If a patient periodically reports location data to the server, then when the frequency of reporting is high enough and the density of patients is low enough, a tracking algorithm can accurately estimate the trajectory of a single patient. Furthermore, if a patient's trajectory goes through sensitive or identifiable places, a patient might see this as private information and these places may also provide con-

nections to the patient’s identity. If there are multiple persons in an assisted living environment, their privacy can be protected if their location data are perturbed and the adversaries cannot distinguish between them through tracking algorithms.

There has been considerable research on location privacy. There are two main categories: (1) using policies to control the release of location data and (2) processing the data in order to hide some parts of users’ trajectories. In this paper, we focus on the second area and do not consider data release policies. One particular technique of interest is the “mix zone method” [3], which creates fixed areas where patients do not report their precise locations to the server. If multiple patients transit through a mix zone simultaneously, the server only records the fact that several people have entered the zone and that several people have left the zone, without associating those entering from those leaving. If a patient stays in the zone long enough, then tracking algorithms will be unable to link a patient’s trajectory entering the zone with one of those leaving. The existing method uses only fixed areas for mix zones and requires that they be pre-configured, and a further limitation is that it is not very effective when there are not many people going through the mix zones at the same time. Another approach to protecting user location privacy is a method proposed by Hoh and Gruteser [13] that perturbs the paths of users. Whenever two users are close enough (their distance is less than a threshold that can be tolerated by supported applications), the paths of the two users will be perturbed and be forced to cross with each other; the idea is that whenever two users’ paths intersect, it is more difficult for a tracking algorithm to follow the real trajectory of a particular user. A privacy metric is presented in [13] that defines privacy in terms of confidence and spatial distance. An algorithm constructed by the authors using a constrained non-linear optimization problem is shown to maximize the proposed metric. The method perturbs users’ paths in order to confuse the tracking algorithm used in the optimization process; however, it operates on two users’ paths at a time. Thus, it is difficult for this method to perturb the location samples of a large number of users efficiently.

In this paper, we propose a method called dynamic mix zones to perturb location information efficiently in order to minimize the chance of it being abused to derive identity information. The dynamic mix zones method can be seen as a way of combining the ideas of mix zones and path confusion by introducing the dynamic creation of mix zones in order to remove the need for planning mix zones in advance while still utilizing the intuitive and practical mix zones to mix patients’ trajectories. Section 2 describes the related work on privacy issues in location samples in more detail. Section 3 gives a formal definition of the problem. Section 4 describes the dynamic mix zones method. Section 5 propose two different kinds of privacy metrics and describes a simulation environment along with the results of experiments using it. Finally, Section 6 summarizes the paper.

## 2. RELATED WORK

There has been considerable research in location privacy beyond the specific contributions mentioned above. Ackerman *et al.* [1] discussed the requirements for the use of wireless location information and developments in the law and regulations governing the use of wireless location information in the United States. Sneekenes [15] elaborated the concept of an observation of a located object and proposed the idea that the individual should be able to adjust the accuracy of their location data. Schilit *et al.* [14] considered privacy risks, economic damages, and network privacy issues affected by the capability of cellular carriers identifying the location of emergency callers using mobile phones. Beresford *et al.* [2, 3]

proposed the mix zone method to enhance user privacy in location-based services. Gruteser *et al.* [10] proposed to “cloak” (degrade) location information; later, this was extended [9] to a middleware architecture and algorithms to adjust the resolution of location information along spatial or temporal dimensions. Gunter *et al.* [11] presented an analog of an access control matrix to model privacy concepts. Hoh and Gruteser [13] reformulated the privacy problem as a constrained optimization problem where selected segments of paths are perturbed.

## 3. PROBLEM DEFINITION

In this paper, we propose to extend the mix zone method to a new dynamic mix zone. Several privacy metrics are proposed and experiments are conducted to evaluate the performance of our method. In this work, we focus on perturbing location samples collected and stored in location servers. Our goal is to process raw location data before it is used by other applications and to preserve patient privacy by making it more difficult for an adversary to estimate the trajectories of patients. We assume that the more possible trajectories an adversary might need to consider a patient associated with, the more privacy is preserved. Assume all the patients report their locations periodically, and the location data stored in location server includes all the locations of every patient at every time stamp. If there are  $n$  persons in the monitored area, let  $M = \{m_1, m_2, \dots, m_n\}$  denote a set of collected locations, where  $m_i$  denotes the location of the  $i$ th person reported and  $M^t$  denotes the set of collected locations reported at time  $t$ .

Assuming the whole time period is  $T$  and the observed locations are the real locations of persons, the  $i$ th person’s trajectory is  $P_i = \{p_i^1, p_i^2, \dots, p_i^T\}$  ( $p_i^t \in M^t, 1 \leq t \leq T$ ). If the data are published without any change, then an adversary can assemble all the location data, and using a tracking algorithm, can track a person throughout the monitored area. Let  $A_i = \{a_i^1, a_i^2, \dots, a_i^T\}$  ( $a_i^t \in M^t, 1 \leq t \leq T$ ) denote the adversary’s hypothesis on the  $i$ th person’s trajectory, whose true trajectory is  $P_i$ . If the adversary can track the whole trajectory of a person without error, which means  $A_i = P_i$ , the privacy of this person can be considered completely compromised, since the whole trajectory is exposed. Assuming when the data is shared among other researchers, their applications’ tolerance on difference between altered locations and original locations is less than a threshold  $\delta$ . Some applications may not need the accurate location samples, and only need statistical information. Let  $C = \{c_1, c_2, \dots, c_n\}$  denote the set of changed locations corresponding to  $M$  and  $C^t$  denote the set of changed locations at time  $t$ . The Euclidean distance between a changed location and its original value  $d = |m_i - c_i| \leq \delta$ . Our goal is to prevent malicious users from compromising patients’ identities while keeping the data usable to other researchers.

## 4. DYNAMIC MIX ZONES

We introduce a concept called the dynamic mix zone. Dynamic mix zones are small areas created dynamically according to the movements of objects. The dynamic mix zone is a virtual zone where location samples are changed or deleted. These mix zones are not generated beforehand or pre-specified. As with mix zones [2], pre-specified mix zones would be hard to configure and sometimes may not work well, since some objects might stay in a mix zone for only a very short time and since the selected mix zones might not turn out to be in the best locations to mix objects.

In this section, we introduce methods for generating dynamic mix zones in order to protect the location privacy of monitored objects. First, a simple two person/two trajectory case is studied. The

method is then extended to allow for multiple persons' trajectories.

#### 4.1 Two Person/Two Trajectory Case

When two persons are walking close to each other, a mix zone is dynamically created, and their separate trajectories are mixed to confuse a possible adversary. When the two persons diverge from each other and changing the location samples can no longer be done within the allowed distortion bounds, the mix zone is demolished and the two trajectories are no longer altered. Figure 1 is an example of this process. Two objects move closer to each other and move apart. When they are close enough at time  $t_1$  and  $d_1 \leq L$ , where  $L$  is a predefined threshold, their location reports are changed by creating new perturbed coordinates. When they are moving apart at time  $t_2$  and  $d_2 > L$ , their location reports are no longer changed. Through changing the location reports, the trajectories between  $t_1$  and  $t_2$  are mixed, and the zone between  $t_1$  and  $t_2$  is a dynamically created mix zone. These two persons' locations are perturbed in the dynamic mix zone to mix their trajectories and protect their privacy. When patients' trajectories are far from each other and the dynamic mix zones can not be created, patients' locations can be perturbed with a random noise less than the application specific limit  $\delta$ , however, the random noise does not help preventing an adversary from tracking a patient if there are no other patients around. If a person moves alone and visits some sensitive locations, the sensitive information can not be protected unless she doesn't reveal her location data when she is at the sensitive location because the random noise added on the location data is limited due to the applications' requirements and an adversary can always use tracking algorithms to track her. In this case, this person can use self-configured policies to regulate whether to report her location data at a specific location. Using policies to protect location privacy is not this paper's focus. We focus on privacy problems raised when location data are published and the privacy is defined as the complete and correct trajectory of a person.

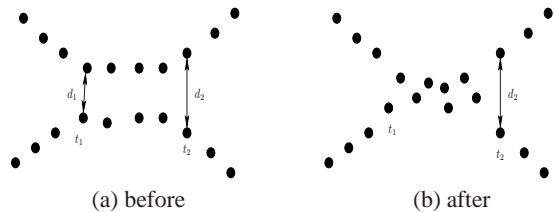


Figure 1: Two persons mix zone

Assuming the original position of an object is  $(x, y)$ , the distance between its new perturbed position  $x', y'$  and the original position should be less than  $\delta$ , which is the toleration threshold of applications. Figure 2 shows an example of generating new positions. The two original positions  $A$  and  $C$  are  $(x_1, y_1)$  and  $(x_2, y_2)$ .  $A'$  and  $C'$  are corresponding new locations for  $A$  and  $C$ . To fully mix  $A$  and  $C$ , which means

$$P(A'|A) = P(C'|A) \text{ and } P(C'|C) = P(A'|C), \quad (1)$$

$A'$  and  $C'$  should be randomly selected from the same area. We can let the new random locations generated only on the line  $AC$ . It works for two persons' case, however, it is hard to extend to multiple persons' case. Thus, we use a circle with a pre-configured radius  $r$  to cover  $A$  and  $C$ . If we have multiple persons, it is easy to generalize the method using a circle to cover multiple persons. We need to define how large the disc should be. Since the maximum change of a location sample should be less than  $\delta$  and in a circle with radius  $\delta/2$ , the maximum distance of any two points is less

than  $\delta$ , we can use circles with radius  $\delta/2$  to cover multiple persons. In other words, we can set the predefined threshold  $L = \delta$  and if the distance of two patients is less than  $\delta$ , their location samples are randomized.

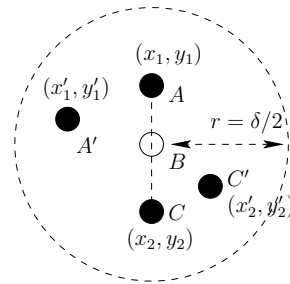


Figure 2: Randomize locations

In Figure 2, let the new positions  $A'(x'_1, y'_1)$  and  $C'(x'_2, y'_2)$  be two random points in the area centered at midpoint  $B$  with radius  $r = \delta/2$ . When  $L = \delta$ , we can see that the reporting of randomized positions fulfills the required application-specific limit  $\delta$  on perturbation. Since  $A'$  and  $C'$  are two uniformly random selected points from the same area,  $P(A'|A) = P(C'|A)$  and  $P(C'|C) = P(A'|C)$ . In this way, the two original points are effectively mixed with each other.

In Figure 1(b), two trajectories are changed according to the above method until their distance  $d_2 > \delta$ . Thus, when the distance between two different person's locations is less than  $\delta$ , or in other words when their locations can be covered by a circle with radius  $\delta/2$ , their locations are mixed in a circle area with radius  $\delta/2$ . From Figure 1, we can see that for the case of two persons' trajectories, this is a very easy way to mix them. In the following section, we will extend this method to mixing multiple trajectories.

#### 4.2 Multiple Person/ Multiple Trajectory Case

Multiple trajectories case is the generalization of two persons' case. The straightforward way is to separate the multiple persons trajectories problem into several problems that includes only two person trajectories, and then use the method for two persons' case to solve them. In [13], the multiple paths problem is decomposed into multiple two path problems because of the high computational cost of their algorithm for altering location samples. The algorithm gets the perturbed location samples from the solution of a constrained non-linear optimization problem, which utilizes the return value of Reid's Multiple Hypothesis Tracking(MHT) algorithm. If the algorithm solve the multiple trajectories case directly, it needs to store  $n^{k-1}$  hypothesis, predict and update the state variables based on these hypothesis, where  $n$  is the number of persons,  $k$  is the number of the time stamps of the whole process. After decomposing multiple paths problem into multiple two path problem, the algorithm still needs to maintain  $2^{k-1}$  hypothesis. Thus, the computation cost increases very fast when the trajectories becomes longer and have more location samples. Moreover, in the case that three persons are very close to each other, and they are far away from other people, only selecting two of them to mix may not be a good solution. So, we need an easy way to mix three or trajectories without high computational cost. In this section, we propose a method for multiple trajectories case, that have a lower computation cost and have a better scalability for multiple trajectories case. Our method for multiple trajectories case is extended from the method for two persons' case and exploit geometric disk covering problem.

In our method for two trajectories, two locations are mixed if they can be covered by a circle with radius  $\delta/2$ . For multiple trajectories, we can also use this idea such that if multiple locations can be covered by a circle with radius  $\delta/2$ , their locations are mixed. Previously we mentioned that if more persons' locations can be covered by a single disc, which corresponds to a dynamically created mix zone, when they move apart, it is more difficult to track one of these persons. For example, if there are two persons in a mix zone, when they leave each other, the probability of successfully tracking one of these two persons is  $1/2$ . If there are five persons in a mix zone, then the probability of successfully tracking one of them is only  $1/5$ . Thus, we want a circle can cover as many persons as possible. In other words, we want to find minimum circles to cover all the persons in order to mix them better. For the location reports at time  $t$ , finding the minimum number of circles to cover them is a geometric disk covering problem.

#### 4.2.1 Geometric disc covering problem

Here we briefly describe the geometric disk covering problem. Given a set of  $n$  points in the plane, find the minimum number of disks of prescribed radius  $r$  to cover them. This problem has been proved to be NP-complete [6]. Polynomial approximation algorithms that provide a suboptimal solution that is within a constant approximation factor of the optimal one have been proposed in [4, 8, 12]. Franceschetti *et al.* [7] considered using a grid as centers of discs and gave an approximation method. The theorem proved by [7] is as follows.

Consider a square lattice where the distance between two neighboring lattice vertices is  $R$ . Call a disc of fixed radius  $r$ , centered at a lattice vertex, a grid disc. The number  $N$  of grid discs that are necessary and sufficient to cover any disc of radius  $r$  placed on the plane is given by

1. CASE 1. For  $r/R < \frac{\sqrt{2}}{2}$ ,  $N$  does not exist.
2. CASE 2. For  $\frac{\sqrt{2}}{2} \leq r/R < \frac{\sqrt{10}}{4}$ ,  $N = 6$ .
3. CASE 3. For  $\frac{\sqrt{10}}{4} \leq r/R < 1$ ,  $N = 5$ .
4. CASE 4. For  $1 \leq r/R < \frac{5\sqrt{2}}{4}$ ,  $N = 4$ .
5. CASE 5. For  $r/R \geq \frac{5\sqrt{2}}{4}$ ,  $N = 3$ .

This theorem tells us that if we only use the discs centered at a lattice vertices to cover the locations of persons, how many discs we need to cover any disc on the plane, in other words, how many more discs we need using grid discs than the optimal solution that can put a disc anywhere. Although we can not find the optimal solution to the geometric disk covering problem, only using the grid discs can still give us a good solution. The advantage of the approximation method is having a short running time.

We will use a heuristic method based on this theorem to create dynamic mix zones. For every time  $t_i$ , an approximate solution to the geometric disk covering is found. The disks can be used as dynamic mix zones, and the location reports are randomized inside each disk. After this process, the data published to the patients is difficult for an adversary to use to trace one single person correctly, but the data is still usable for the applications.

#### 4.2.2 Heuristic dynamic mix zone generation

The data reported are in the form of  $\{M^1, M^2, \dots, M^T\}$ . Every  $M^t$  is the set of locations of all the persons. Assuming persons are in motion, a mix zone is created dynamically whenever several persons are close to each other; when they later move apart, this

mix zone is demolished. We first study the creation of a dynamic mix zone at the first time objects are close enough to trigger one.

Based on the result of [7], a grid can be used as the centers of geometric discs. The basic idea of the heuristic dynamic mix zone generation method is to put a square lattice in the field: the possible mix zones are then the circles centered on the grid points. On every point on the grid, a disc with radius  $\delta/2$  is placed and checks how many locations are in its vicinity. If a disc can cover more than one object position, then these positions can be mixed in this disc, and this disc is a mix zone for these location samples. The distance between two neighboring lattice vertices depends on the approximation factor needed. From the theorem of [7], we can get the following result straightforwardly.

LEMMA 1. *Using discs of fixed radius  $\delta/2$  centered at a lattice vertex to cover locations, the number  $N$  of these discs that are necessary and sufficient to cover any disc of radius  $\delta/2$  placed on the plane is given by:*

1. CASE 1. For  $R > \frac{\sqrt{2}}{2}\delta$ ,  $N$  does not exist.
2. CASE 2. For  $\frac{\sqrt{10}}{5}\delta < R \leq \frac{\sqrt{2}}{2}\delta$ ,  $N = 6$ .
3. CASE 3. For  $\frac{1}{2}\delta < R \leq \frac{\sqrt{10}}{5}\delta$ ,  $N = 5$ .
4. CASE 4. For  $\frac{\sqrt{2}}{5}\delta < R \leq \frac{1}{2}\delta$ ,  $N = 4$ .
5. CASE 5. For  $R \leq \frac{\sqrt{2}}{5}\delta$ ,  $N = 3$ .

To minimize the number of discs in the approximate solution, we use  $R = \frac{\sqrt{2}}{5}\delta$  as the distance between two neighboring lattice vertices.

Thus, the dynamic mix zone generation method for a single time is as follows.

1. First, a lattice is put in the area of location samples, and the distance between two neighboring vertices is  $\frac{\sqrt{2}}{5}\delta$ .
2. Second, for every lattice vertex, a disc with radius  $\delta/2$  is placed on that vertex if the disc can cover more than one location. However, a location can only be covered at most once, so a greedy method is used to find the final disc covering. The disc that covers the largest number of locations is repeatedly found and added to the solution until there is no disc that can cover more than one location left.
3. Third, for every disc in the solution, the location samples in it are randomized using the method described in Section 4.1. For each position, a randomly selected location is chosen uniformly from the disc to replace it.

To allow for location data over multiple time points, we can simply reapply the above method.  $M^1, M^2, \dots, M^T$  can be viewed independently, and the single time method can be applied to every  $M^t$  separately.

## 5. EXPERIMENTS

In this section, we first propose two different privacy metrics and then use experiments to evaluate the dynamic mix zones method.

### 5.1 Privacy Metrics

To study the performance of the location perturbation methods, we need some privacy metrics. We can not use a single privacy metric to cover all the aspects of privacy preserved. Thus, in this section, two different privacy metrics are introduced to evaluate different aspects of the privacy preserved.

### 5.1.1 Indistinguishability property

The goal of location privacy is to prevent an adversary from tracking the whole trajectories of a patient, which means after mixing, a patient's trajectory can not be identified from all the paths in the location data. In other words, the more other persons a person has been mixed with, the better privacy she can get. The locations of persons covered by a disc are randomized in order to make them indistinguishable. We assume that if two objects had been in one mix zone together, their identity are now undistinguishable. In other words, even after they leave the mix zone, an adversary can not distinguish between them without further information. Thus, for a single person, if she has been with more persons in any mix zone, she is undistinguishable with more persons. We introduce a metric called indistinguishability to quantify the performance of our methods on this aspect.

The more objects are combined, the more privacy is preserved. Figure 3 is an example. Each grid point is a possible mix zone center. If the circle centered at A is assigned as a mix zone, it can cover  $m_1, m_2, m_3$ , and  $m_4$ . The mix zone centered at B can cover  $m_2, m_3, m_4, m_5, m_7$ , and  $m_8$ . The mix zone centered at C can cover  $m_5, m_6, m_7$ , and  $m_8$ . So, we have two choices: using A and C to cover all the locations, or using B to cover all the locations except  $m_1$  and  $m_7$ . If A and C are selected,  $m_1, m_2, m_3$ , and  $m_4$  are mixed and  $m_5, m_6, m_7$ , and  $m_8$  are mixed. If B is selected, all the points except  $m_1$  and  $m_7$  are mixed. A and C can mix  $N_u = \binom{4}{2} + \binom{4}{2} = 12$  pairs of locations, and B can mix  $N_u = \binom{6}{2} = 15$  pairs of locations. Thus, different choices of covering discs result in different mix zones and the number of pair locations being mixed is different. We use the number of pairs of undistinguishable objects  $N_u$  as the metric to evaluate the performance of the dynamic mix zones approach. We call this the **indistinguishability**. For multiple objects and multiple time points,  $N_u$  is the number of pairs of undistinguishable objects in the whole process.

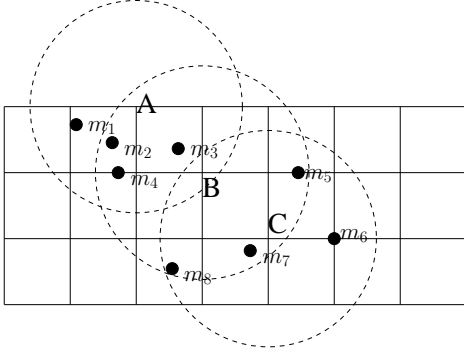


Figure 3: Heuristic dynamic mixing zone for a single time point

The indistinguishability ratio  $r_u$  is defined as the ratio of the indistinguishability value  $N_u$  over the number of all the pairs of locations  $N$ :

$$r_u = \frac{N_u}{N}$$

. The indistinguishability ratio measures how well the algorithm mix all the objects.

If the distance between two objects was less than the threshold  $\delta$  in a time point, they are a mixable pair. The performance ratio  $r_p$  is defined as the ratio of indistinguishability value  $N_u$  over the

number of all the mixable pairs of locations  $N_m$ :

$$r_p = \frac{N_u}{N_m}$$

. Using this performance ratio, different algorithms can be compared to evaluate how well the algorithm mixes all the potentially mixable objects.

### 5.1.2 Anti-tracking performance

If different data altering algorithms are used, their effects on adversaries' tracking algorithms may be different. Thus, we need a metric to compare the effect of various privacy protection methods on adversaries' tracking methods. The distance error has been used in previous research for this purpose; the distance error is the sum of the differences between correct assignments and incorrect assignments. However, sometimes the distance error may not truly show how much privacy has been preserved. Consider an adversary tracking one person whose true trajectory consists of 100 locations. When algorithm A is used to change the locations, the adversary can get most of them right, but a few of them have a large distance error. When another algorithm B is used, the adversary assign most of the locations wrong, but the distance error sum over all the locations is less than A. In this case, it seems intuitively clear that using B can preserve more privacy than A, despite the fact that it achieves a worse score on this metric.

Alternatively, we can consider using the number of locations that an adversary assigns to the wrong person as a privacy metric. However, this may also not truly reflect the true amount of privacy preserved. Consider two tracks that are always close to each other; when reported locations are changed, an attacker may be confused between this pair of tracks, but the distance between the incorrect track and the true track is always small—thus, true privacy is not well preserved. We introduce a new metric called anti-tracking performance as follows to address the limitations of the preceding measures.

**DEFINITION 1.** Assume a tracking algorithm  $S$  determines an object  $A$ 's track is  $Tr_A = \{a^1, a^2, \dots, a^T\}$ , and the object's original track before location alteration using algorithm  $L$  is  $B = \{b^1, b^2, \dots, b^T\}$ . The **anti-tracking performance** of  $L$  against  $S$  for  $A$  is  $P_A = \sum_{i=1}^T f(a^i, b^i)/T$ , where

$$f(x, y) = \begin{cases} 1 & \text{if } D(x, y) > \delta \\ 0 & \text{if } D(x, y) \leq \delta \end{cases}$$

$D(x, y)$  is the distance between  $x$  and  $y$ ,  $\delta$  is the threshold defined by the application.

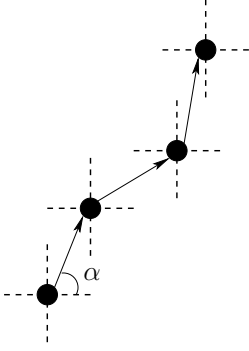
The anti-tracking performance of  $L$  against  $S$  for multiple objects throughout a period of time is

$$P = \frac{1}{n} \sum_{j=1}^n P_j = \frac{1}{nT} \sum_{j=1}^n \sum_{i=1}^T f(a_j^i, b_j^i),$$

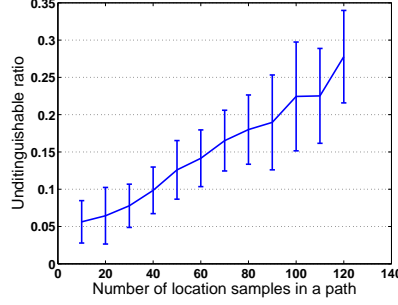
where  $n$  is the number of objects,  $a_j^i$  denotes the object  $j$ 's location at time  $i$  computed by the attacker, and  $b_j^i$  denotes object  $j$ 's real location at time  $i$ .

## 5.2 Datasets

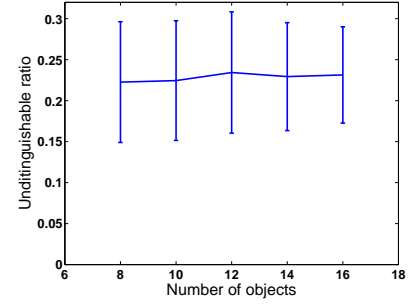
The location data for multiple objects over multiple time points were randomly generated as follows. There are 10 objects moving in an area of  $100 \times 100$  meters. The application-specific threshold  $l$  is set to 10 meters. The locations of objects are reported every 10 seconds, and an object moves 1 meter in a random direction at



**Figure 4: Moving path of an object with an assigned moving trend**



**Figure 5: Indistinguishability ratio with different path lengths**



**Figure 6: Indistinguishability ratio with different numbers of objects**

every step. In order to prevent the random path of an object consisting of circles centered at the initial position, a moving trend is assigned to each objects in the beginning. The trends can be categorized into four classes by dividing 360 degrees into four 90 degrees: northeastern, northwestern, southwestern, and southeastern. Each object is assigned one of these categories in the beginning, and the object’s random moving direction at each step is generated based on that category. For example, if an object’s moving trend is northwestern, the moving direction of this object’s every step is  $D = R + 90$ , where  $D$  is the angle between this object’s moving direction and east,  $R$  is uniformly random generated between 0–90 degrees. Figure 4 is an example of one object’s moving path with a northeastern moving trend.

### 5.3 Indistinguishability ratio and performance ratio

To evaluate performance in terms of the indistinguishability property for the heuristic dynamic mix zones method, random trajectories of 10 objects lasting different lengths of time were generated and the heuristic dynamic mix zones method was applied to the locations. Figure 5 shows the indistinguishability ratio of the objects using the heuristic dynamic mix zones method. It is clear that the indistinguishability ratio increases with an increase in the number of location samples of objects. In other words, objects are more indistinguishable when they travel for a longer time. When objects travel for a longer time, it is plausible that there are more object encounters, each resulting in additional chances to be covered by a dynamic mix zone. Figure 6 shows that with different numbers of objects in the area, the indistinguishability ratio doesn’t change much.

We also computed the performance ratio of the heuristic dynamic mix zones method compared to the optimal solution (Figure 7). The result shows that the heuristic dynamic mix zones method can get a performance ratio OF around 0.9 all the time, which means that almost all of mixable objects have been mixed if they have been close enough. Figure 8 shows the performance ratio with different numbers of objects in the area; the performance is almost the same when number of objects changes.

### 5.4 Anti-tracking performance

To evaluate the performance of our heuristic dynamic mix zones method against an attacker’s tracking method, we can use existing tracking algorithms to track objects using the location samples that have been processed by our dynamic mix zones method. In our current experiments, we have used a simple tracking algorithm based on a Kalman filter. Assume that the movement of objects can be

described using the linear Gaussian state space model as

$$\begin{aligned} y_t &= Z_t \alpha_t + \epsilon_t, & \epsilon_t &\sim N(0, H_t), \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t), \quad t = 1, \dots, n, \end{aligned}$$

where  $y_t$  is a vector of observations and  $\alpha_t$  is an unobserved state vector. The system matrix

$$T_t = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and the observation matrix

$$Z_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The Kalman filter is as follows,

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\ K_t &= T_t P_t Z_t' F_t^{-1}, & L_t &= T_t - K_t Z_t, \\ a_{t+1} &= T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t L_t' + R_t Q_t R_t'. \end{aligned}$$

Let  $Y_t = \{y_1, \dots, y_t\}$ , then the log-likelihood of a series of observations

$$\begin{aligned} \log L(y) &= \sum_{t=1}^n \log p(y_t | Y_{t-1}) \\ &= -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n (\log |F_t| + v_t' F_t^{-1} v_t). \end{aligned}$$

The log-likelihood can easily be computed from the output of the Kalman filter [5]. The tracking algorithm for an object works as follows.

1. An object  $i$ ’s trajectory starts from the first location sample  $m_i^1$ .
2. Let  $P_i = p_i^1, \dots, p_i^t$  denote the current trajectory of object  $i$ . After receiving the next set of location data  $M^{t+1}$ , compute the log-likelihood of  $\{p_i^1, \dots, p_i^t, m_x^{t+1}\}$ ,  $x \in (1, \dots, n)$ .
3. Let  $p_i^{t+1} = m_{max}^{t+1}$ , which gets the maximum log-likelihood in step 2.
4. Go to step 2 until the there is no new location data.

Using this simple tracking method, we can track an object using location data that have been preprocessed by the dynamic mix zones method. After we get the trajectory of an object, we compare it with the original location data according to the anti-tracking

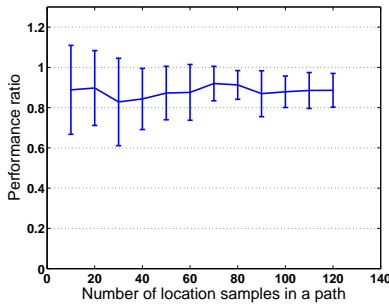


Figure 7: Performance ratio with different path lengths

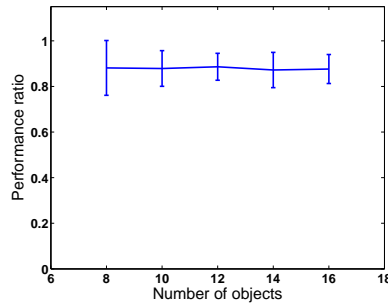


Figure 8: Performance ratio with different number of objects

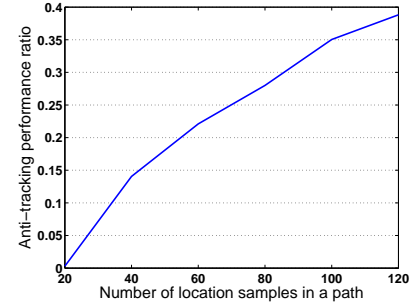


Figure 9: Anti-tracking performance with different path lengths

performance metric in Section 5.1.2. Figure 9 shows the result of our experiments. When the travel time of objects is longer, anti-tracking performance increases. This is because more objects are mixed, and an attacker is more easily led in wrong directions. We

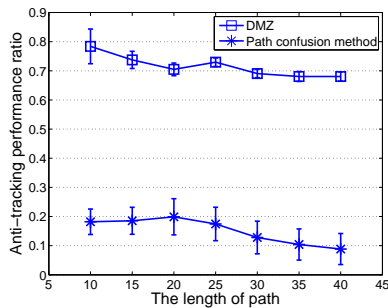


Figure 10: Anti-tracking performance comparison

compared our method with path confusion method in [13]. To compare with their solutions, we use the same setups as theirs as follows. The paths of objects are no longer limited in a specific area. The application specific threshold is set to 150 meters. Figure 10 shows the results for 5 objects. We can see that the performance of DMZ is better than path confusion method. The reason maybe due to that in their method, perturbation is only done on two paths at a time in a segment. In our method, every time we mix all the location samples that are close enough, which can randomize multiple objects at the same time.

## 6. CONCLUSIONS AND FUTURE WORK

We discussed the problem of preserving privacy for patients' location data in assisted living environments. A whole trajectory of a patient being monitored may leak private information. We proposed a method called dynamic mix zones to mix the locations of monitored objects based on their relative locations instead of using fixed mix zones. Two privacy metrics are proposed to evaluate its performance. Experiments shows that the longer objects travel, the better their privacy is preserved when using the dynamic mix zones method. Compared with previous methods, since dynamic mix zones method randomizes multiple objects at the same time, it has better performance on mixing location samples to prevent malicious tracking. For future work, how to mix the location samples on the sensors when they are being collected might be a possible direction.

## 7. REFERENCES

- [1] L. Ackerman, J. Kempf, and T. Miki. Wireless location privacy: A report on law and policy in the united states, the european union, and japan. *DoCoMo USA Labs Technical Report DCL-TR2003-001*, 2003.
- [2] A. R. Beresford and F. Stajano. Location privacy in pervasive computing. *IEEE Pervasive Computing*, 2(1):46–55, 2003.
- [3] A. R. Beresford and F. Stajano. Mix zones: User privacy in location-aware services. In *Second IEEE Annual Conference on Pervasive Computing and Communications Workshops*, page 127. IEEE, March 2004.
- [4] H. Bronnimann and M. T. Goodrich. Almost optimal set covers in finite VC-dimension: (preliminary version). In *SCG '94: Proceedings of the tenth annual symposium on Computational geometry*, pages 293–302, 1994.
- [5] J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press Inc., New York, 2001.
- [6] R. J. Fowler, M. S. Paterson, and S. L. Tanimoto. Optimal packing and covering in the plane are NP-complete. *Information Processing Letters*, 12(3):133–137, June 1981.
- [7] M. Franceschetti, M. Cook, and J. Bruck. A geometric theorem for network design. *IEEE Transactions on Computers*, 53(4):483–489, 2004.
- [8] T. Gonzales. Covering a set of points in multidimensional space. *Information Processing Letters*, 40(4):181–188, 1991.
- [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)*. USENIX, 2003.
- [10] M. Gruteser, G. Schelle, A. Jain, R. Han, and D. Grunwald. Privacy-aware location sensor networks. In *HotOS IX: 9th USENIX Workshop on Hot Topics in Operating Systems*, 2003.
- [11] C. A. Gunter, M. J. May, and S. Stubblebine. A formal privacy system and its application to location based services. In *Privacy Enhancing Technologies (PET)*, 2004.
- [12] D. Hochbaum and W. Maass. Approximation schemes for covering and packing problems in image processing and VLSI. *Journal of ACM*, 32(1):130–136, 1985.
- [13] B. Hoh and M. Gruteser. Protecting location privacy through path confusion. In *First International Conference on Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005.*, 2005.
- [14] B. Schilit, J. Hong, and M. Gruteser. Wireless location

privacy protection. *IEEE Computer*, pages 135 – 137, December 2003.

- [15] E. Snekkenes. Concepts for personal location privacy policies. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 48–57. ACM Press, 2001.