Dummy Trajectory Generation Scheme Based on Deep Learning

Conference Paper · November 2019

CITATIONS
0

READS
218

3 authors, including:

Yining Liu
Guilin University of Electronic Technology
92 PUBLICATIONS **1,479** CITATIONS

SEE PROFILE

Dummy Trajectory Generation Scheme Based on Deep Learning

Jiaji Pan[1,2], Jingkang Yang[1] and Yining Liu[1](✉)

[1] School of Computer and Information Security, Guilin University of Electronic Technology, China
ynliu@guet.edu.cn

[2] College of Computer Science and Technology, HenyangNormal University, Henyang 421002, China

Abstract. Nowadays, the traditional dummy trajectory generation algorithm used to protect users' trajectory privacy usually uses statistical methods to build trajectory model. Because the human mobility model is a complex equation, it is difficult to use mathematical methods to model, so the established trajectory model can not consider the human mobility model which restricts the formation of trajectory. Therefore, traditional dummy trajectory generation algorithms can not defend against data mining attacks based on in-depth learning. In this paper, LSTM (Long Short-Term Memory) is used to design the real and dummy trajectory discriminator. Experiments show that data mining based on deep learning can eliminate more than 95% of the algorithm generated trajectories, and the error rate of real trajectory is less than 10%. We restrict the traditional dummy trajectory generation algorithm to human mobility model, and design a dummy trajectory generation strategy so that the generated trajectory can defend against multiple recognition attacks.

Key words: dummy trajectory, human mobility model, LSTM

1 Introduction

In today's society, LBS (Location based services) [1-7] has developed rapidly and is widely used in smart mobile terminals. The LBS service provider and the customer use the trajectory data as the information entity and the Internet as the information carrier for information interaction. A large amount of trajectory data is generated during the interaction process. These trajectory data have rich time and space information, and information about the user's personal interests and social status can be obtained through data mining. However, the trajectory information contains too much data on personal privacy and therefore cannot be published directly. Therefore, how to protect the privacy of users in data distribution and LBS can make the trajectory data play a role and is concerned by researchers. In recent years, researchers have also applied the proposed new technology to data privacy and trajectory privacy protection in LBS. However, the algorithms proposed by current researchers can only resist traditional data mining attacks, and can not resist data mining attacks based on deep learning.

Aiming at the problem of trajectory privacy protection of intelligent mobile terminals, a large number of solutions have emerged in the academic world, which can be roughly divided into two categories: a trusted third-party server-based method and a mobile terminal-based distributed method. These schemes are mainly implemented by means of spatial cloaking, mix-zone, path confusion, and dummy trajectory.

Among them, the dummy trajectory method is most practical because it does not need a trusted third party, can defend against channel-side attacks, and its operation is simple and does not need to cooperate with other users.

However, current researchers have been unable to evade the problems of what trajectory conforms to the real trajectory when designing the dummy trajectory generation algorithm. Most researchers use mathematical modeling methods to try to explain whether the trajectory conforms to the real trajectory characteristics from the perspective of probability theory and statistics and has achieved a series of results. However, it is always impossible to explain the constraints of the formation of the trajectory.

In order to verify that there are some constraints in the process of trajectory formation that can not be explained in the current paper, we design and train a LSTM-based trajectory discriminator in the third part of this paper. We put the dummy trajectory generated by the algorithm ADTGA[10], MLN[8] and MN[8] into the trained LSTM trajectory discriminator, respectively, to prove that the dummy trajectory generated by the classical algorithm can not resist trajectory identification attack based on LSTM. The trajectory discrimination attack proves that the dummy trajectory generated by the classical algorithm and the real trajectory are not distributed in the same space-time, and proves that the dummy trajectory generated by the classical algorithm does not conform to the real human mobility model. In the next part, a complete scheme of dummy trajectory generation is designed.

Contributions. In this paper, we:

- define trajectory points with four-dimensional vectors and describe the space-time distribution of trajectory sections composed of n points with $n \times 4$ matrix.
- design a trajectory detector based on LSTM, which can detect most of the generated trajectories of the algorithm, but the error rate of the real trajectory is very low.
- combine background information, human movement mode and classical dummy trajectory generation algorithm to design a complete real-time continuous trajectory privacy protection scheme.

The rest of the paper is organized as follows. Section 2 introduces Related Works, LSTM-based trajectory discriminator. The LSTM-based trajectory discriminator described in Section 3. We propose dummy trajectory generation strategy in Section 4. Section 5 concludes the paper.

2 Related Works

Kido et al. [8,9] proposed that users could use dummy locations generated by two requests to form dummy trajectories, which pioneered the use of dummy trajectories to protect trajectory privacy. Wu et al. [10] proposed that virtual trajectories could be generated by adding disturbance, taking into account the distance between real trajectories and dummy trajectories and the distance between dummy trajectories, so that the final generated trajectory set could meet the needs of user privacy protection. Ray et al. In order to protect user's trajectory privacy, a method of adding intersection points to the trajectory after rotation is proposed. Kato R et al. [11] assumes that user movement is known beforehand, and proposes a dummy anonymity method based on user movement, in which dummy objects move naturally when multiple locations stop. Niu et al. [12] fully considered the probability information of users sending requests at each location, and formalized the background information using information entropy. A dummy location generation algorithm based on background information was proposed to ensure that the generated dummy location could effectively confuse opponents.

In our paper, in order to verify that the dummy trajectories generated by the traditional dummy trajectory generation scheme do not conform to the human mobility model, we select three algorithms of MN, MLN and ADTGA in the traditional dummy trajectory generation algorithm to generate dummy trajectories, and use our LSTM-based dummy trajectory discriminator to identify these generated dummy trajectories.

3 Dummy Trajectory Recognition Scheme Based on LSTM

Traditional dummy trajectory generation schemes are implemented by mathematical modeling. These algorithms attempt to generate dummy trajectories with the same distribution as real trajectories. In order to prove that the trajectory model established by traditional mathematical modeling methods can not describe the spatial and temporal distribution of real trajectories, we have done the following work to verify our conclusions.

3.1 Dummy Trajectory Identification Scheme

We define a series of trajectory points to form a trajectory section, and two adjacent points form a trajectory segment.

Suppose our detector can detect the trajectory section from m to n points. When a trajectory is detected, it is firstly divided into sections of equal trajectory points(points from m to n). If the number of points in the last part is less than m, it is discarded. Otherwise, use detector to detect.

The dummy trajectory identification scheme is shown in Figure 1. We divide a trajectory into n trajectory sections. Each trajectory section is detected by a trajectory detector, and n results are obtained. It is assumed that k of the n results are dummy. If k/n is greater than the set threshold, the trajectory is judged to be dummy or real.

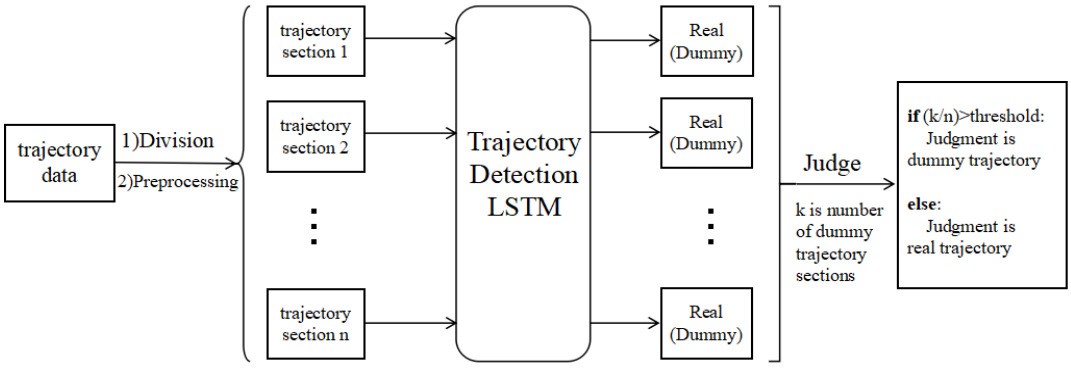


Fig. 1. Dummy trajectory detection model framework based on LSTM

3.2 LSTM Trajectory Discriminator Framework

The LSTM-based trajectory discriminator framework is shown in Figure 2.

- ***Input Layer:*** First, we transform the trajectory from longitude and latitude to plane rectangular coordinate system, then divide the trajectory into equal trajectory sections, and finally preprocess the trajectory points into time series matrix as shown in Table 1. The trajectory set includes real trajectory and algorithm generated trajectory.
- ***Network Layer:*** The first position state of the trajectory section is input into the LSTM network in time sequence, and the final result is

  transmitted to the output layer through the fully connected layer.

- ***Output Layer:*** Softmax operation is performed on the output of the fully connected layer to get the classification results.
- ***Network Training Layer:*** In the network training stage, the label of the trajectory section (true or dummy) and the output of the output layer are combined to calculate the loss value. The weight matrix and bias of the network layer are updated by gradient descent method to reduce the loss value, so that the prediction results are closer to the real results

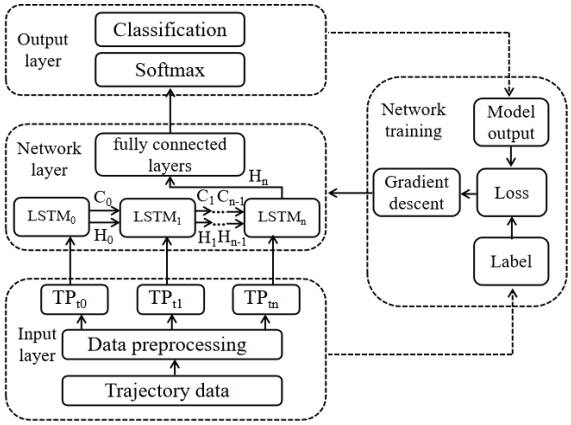


Fig. 2. LSTM-based trajectory discriminator framework

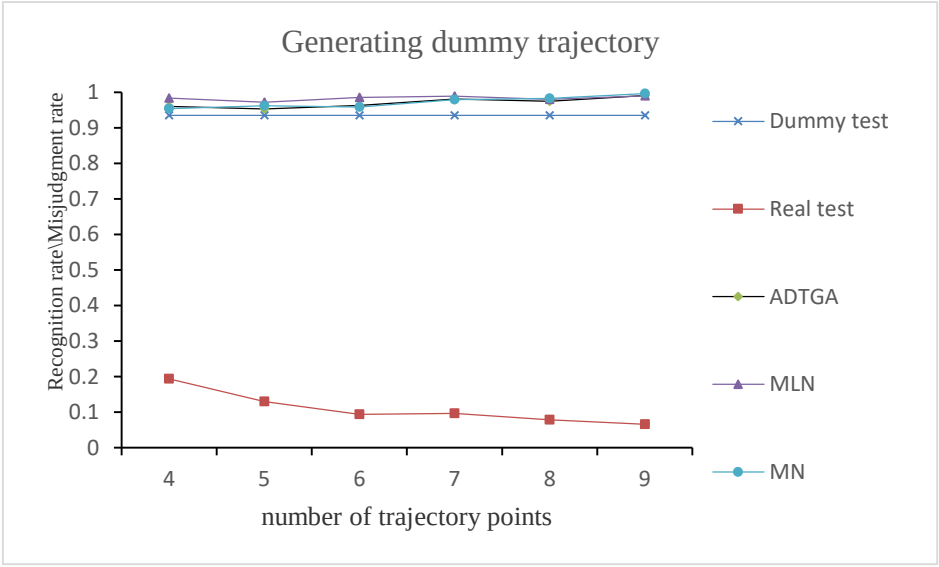


Fig. 3. Detection performance of LSTM-based dummy trajectory detector

Figure 3 shows the detection performance of LSTM-based dummy trajectory detector. On the whole, the average detection rate of this detector is over 95% for the dummy trajectory generated by the classical algorithm, and about 10% for the real trajectory. With the increase of the number of trajectory points in each trajectory section, the detection rate of the generated trajectory increases slightly, and the error rate of the real trajectory decreases greatly.

Experiments show that the dummy trajectories based on LSTM designed by us can detect most of the generated trajectories which do not conform to the real trajectories's patial and temporal distribution, while for the real trajectories, the error rate is low. With the increase of trajectory points, the recognition rate of the generated trajectory is slightly improved, while for the real trajectory, the error rate is significantly reduced, that is, the more the trajectory

points are sampled, the better the detection effect is. When the number of trajectory points reaches 9, the error rate of real trajectory is only 6.6%.

4 Improved Dummy trajectory generation scheme

4.1 Consider three constraints of dummy trajectory generation (priority from high to low):

1) External background constraints, such as inaccessible areas such as waters, green areas, etc.

   Priority: High

   Adversary can easily grasp user's background information (map information). Trajectories that do not conform to background information are easily excluded.

2) Trajectory's own characteristic limitation (human mobility model):

   Trajectory is a broken line segment which is limited to a certain space-time distribution space. (Trajectory features learned by deep neural networks are complex and exist in trajectory. They are difficult to express by equation and can only be expressed by black box of neural networks.)

   Priority: Medium

   Experiments show that the inherent characteristics of the trajectory determine that the dummy trajectory can not be constructed arbitrarily, otherwise it can be easily identified by the dummy trajectory detector based on depth learning. in our previous paper[13], we used the detector based on convolutional neural network and achieved good detection results. Of course, in order to avoid obstacles, people will take some different trajectories from the conventional ones under the background restriction, so the priority ranks behind the external background.

3) Inter-trajectory limitation, space limitation of trajectory point distribution and some other limitations. (Trajectory features modeled by mathematical methods such as probability and statistics are simple and easy to be expressed by equation. Researchers construct trajectory features based on experience, and the description of features is not necessarily accurate or irrelevant to trajectory.)

   Priority: Low

   Background information must be given priority, and the inherent attributes of the trajectory parsed by the deep learning classifier are more reliable than the trajectory characteristics based on empirical mathematical modeling, because the ability of deep learning to obtain deep information is stronger than that of mathematical modeling.

Of course, so far, researchers have designed many excellent pseudo-trajectory generation algorithms. Our dummy trajectory generation strategy combines these excellent dummy trajectory generation algorithms to construct dummy trajectories on the basis of considering background information and inherent attributes of the trajectory itself, so that the generated dummy trajectories can mix the spurious with the genuine.

4.2 Precondition for Constructing Reasonable Dummy Trajectories

The constraints of generating trajectory explain the internal and external causes of forming trajectory.

Internal causes: Human movement model

External Cause: Background information, as long as there are obstacles on the map, the trajectory must be avoided.

Also, when generating multiple dummy trajectories, the dispersion of positions between trajectories should be considered to reduce the leakage probability.

In confronting an adversary, we assume that the adversary has all the information, namely:

1) The adversary has the background information of the location when user requests the LBS service.
2) Adversary trained classifier with deep learning tool to recognize real trajectory and arithmetic generated trajectory
3) The opponent has some other methods to distinguish dummy trajectories, and some other methods proposed in related papers.

Therefore, in order to confront the adversary who knows all the information, the generated dummy trajectory is restricted by these three conditions. Considering these constraints, the dummy trajectory scheme needs the following preconditions.

Background information

. Our scheme uses simplified background information design. we divide the background into three levels as shown in figure 4:

1) Hot region: $\text{Pedestrian flow} > \alpha$
   Hot region is represented by green grid, which represent densely populated areas. When trajectory generation, trajectory points are selected first.
2) Ordinary region: $\alpha \geq \text{Pedestrian flow} > \beta$
   Ordinary region is represented by yellow grid, which represents the relatively sparse region of pedestrian flow, and is the secondary area of trajectory points when trajectory generation.
3) Forbidden region: $\text{Pedestrian flow} \leq \beta$
   Forbidden region is represented by red grid, which represent areas with little pedestrian flow. Trajectory points should avoid these areas when generating trajectories.

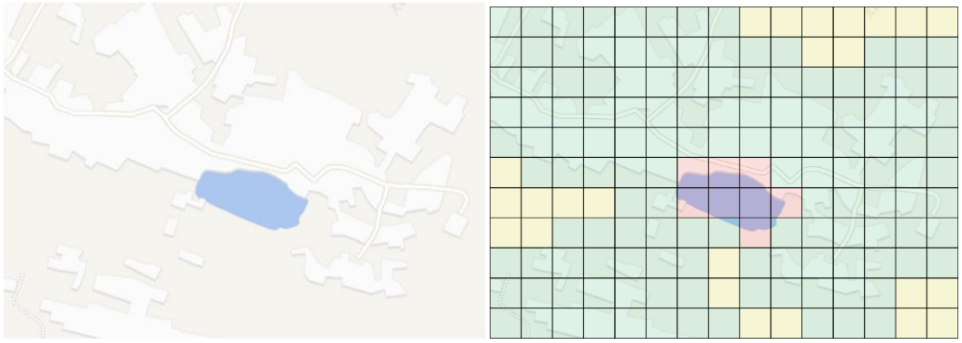


Fig. 4. Map background information

Appropriate trajectory discriminator

. In order to make the generated dummy trajectories conform to the human mobility model, it is necessary to design a reasonable trajectory discriminator to distinguish the real trajectory from the generated trajectory which is not in the same distribution space as the real trajectory, so as to screen the generated trajectory which is the same distribution as the real trajectory.

Experiments show that our LSTM-based trajectory discriminator can recognize more than 95% of dummy trajectories, and the error rate for real trajectories is less than 10%. It proves that our designed trajectory discriminator has the ability to distinguish whether a trajectory is in the same distribution as the real trajectory.

Dummy trajectory generation strategies for some other attack methods

. In the work done by previous researchers, various dummy trajectory generation strategies for enemy attacks have been proposed, and remarkable results have been achieved. These mature strategies against opponent's dummy trajectory prediction can be combined with the former two to design a set of perfect dummy trajectory generation strategies.

4.3 Dummy Trajectory Generation Scheme

Our scheme is to establish the generation of each dummy trajectory point under three restrictions. Under our scheme, we can ensure that user Alice is free from trajectory privacy theft.

Alice travels to a tourist village. Starting from place A of the tourist village, she wants to know some interest points (restaurants, scenic spots, accommodation points, etc.). So she opens an APP to inquire about nearby interest points, but she does not want to reveal her trajectory privacy. At this time, she faces two potential trajectory privacy thieves (assuming that the stealer has the ability to analyze the trajectory and master the background information of the trajectory):

1) Bob intercepting the service request sent by Alice on the channel side
2) LBS Provider (an APP)

So she preprocessed the request on her mobile phone and sent the generated k-1 dummy trajectory points along with the real trajectory points to the LBS provider.

The trajectory point processing process is as follows (generating a dummy trajectory):

Step1: Initialize $k-1$ dummy trajectory starting point at t_0, so that it falls in the green area within a certain distance from Alice. And do the following.

Operation1: Alice began to send service requests for k locations to LBS service providers. Only Alice's mobile terminal knows the authenticity of k locations. The service requests returned by the LBS service provider to Alice's mobile terminal are returned to klocations. Alice's mobile terminal filters out $k-1$ dummy requests and returns the results returned by the LBS server to Alice.

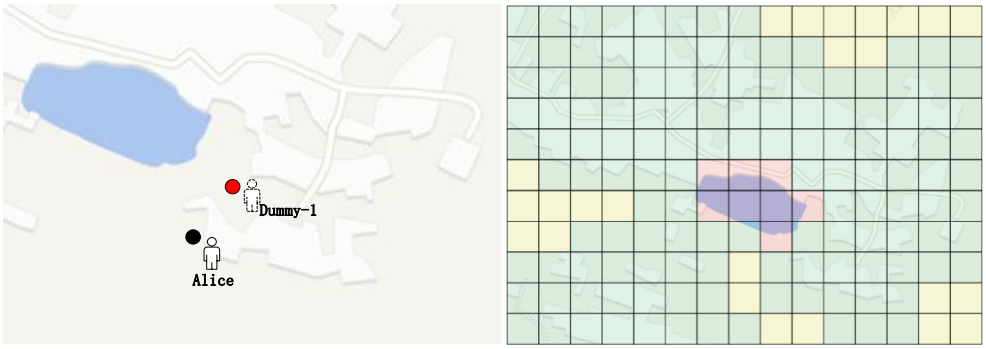


Fig. 5. Step 1 of Dummy trajectory generation strategy

Step2: At t_1 and t_2, since the designed LSTM needs at least three locations to determine whether the trajectory conforms to the real trajectory distribution, the traditional dummy generation algorithm is used to generate dummy points (here we use the classical dummy trajectory generation algorithm MLN). The restriction is that the generated dummy trajectory points must fall in the green or yellow region.

If the feasible region of the dummy trajectory points generated by MLN algorithm does not intersect with the feasible region of the background, the dummy trajectory points that meet the background constraints are selected first, and the trajectory points are as close as possible to the feasible region of the dummy trajectory points generated by MLN algorithm.

Do Operation1.

Step3: From time t_3 to the end of the request service, Alice's mobile terminal does the following:

A. According to the points of the previous moment, the range of the dummy position points at the current moment is determined and divided into grids. The priority of each grid's landing point is identified by different colors, the red area represents the non-landing point, and the green area represents the priority selection area.

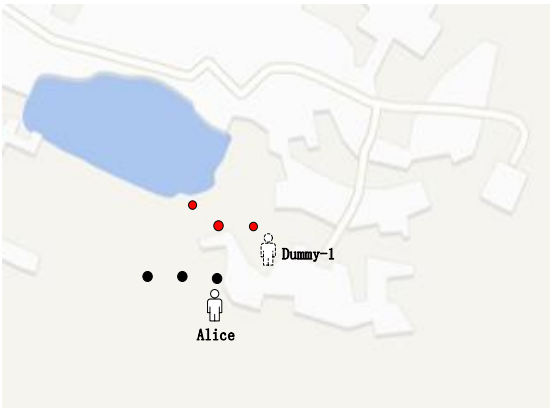


Fig. 6. Step 2 of Dummy trajectory generation strategy

B. Traverse all green and yellow grids. For each grid, assume that the midpoint of the grid is a dummy feasible point, and input it into the trained LSTM network together with the dummy location points of previous moments. If the network judges real, the point is a feasible point, otherwise it is an infeasible point, and the feasible point is a hollow Pentagon as shown in Step 3.B.

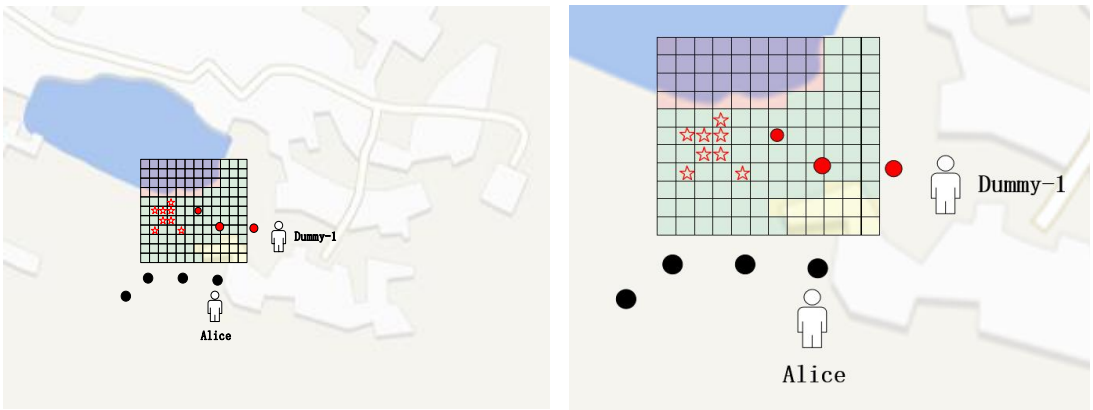
C. Determine the feasible region of the dummy trajectory points according to the MLN dummy trajectory generation algorithm, such as the dotted line frame area shown in Step 3.C.

D. After determining the feasible region of the dummy trajectory points, the selection of the dummy trajectory points follows three restrictive priorities, namely:

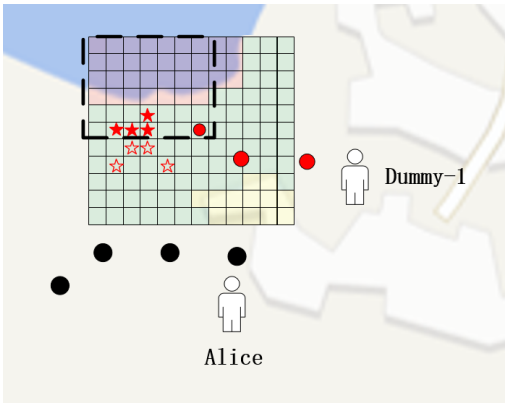
1) If there are hollow pentagonal stars in the dotted frame, these points are selected as the set of points to be selected for dummy trajectory, and one point in the set of points is selected randomly as the set of dummy trajectory points. The set of points to be selected is shown by the solid Pentagon in the figure.

2) If there is no pentagonal star in the dotted frame, look for the pentagonal star outside the dotted frame near the dotted frame as the dummy trajectory point.

3) If there is no pentagonal star in the background feasible region, the random points in the dashed frame are selected, such as the intersection of the dashed frame and the background feasible region, and the points in the background feasible region are selected as the dummy trajectory points.



Step 3.A of Dummy trajectory generation strategy Step 3.B of Dummy trajectory generation strategy



Step 3.C of Dummy trajectory generation strategy.

Fig. 7. Step 3 of Dummy trajectory generation strategy

Algorithm 1: Group t Dummy locations Generation Algorithm

```
For i in range (k-1):
  Determining feasible grids based on background information
  Generated = 0
  If  (t <= 3):
        Generating candidate dummy trajectory point set Using Classical Algorithms
        If (feasible grid and candidate dummy trajectory point set intersect):
              Random selection of intersection points
              Generated = 1
  Else:
        Determining feasible points HMMPs (Points conforming to Human Mobility Model) using LSTM
        If (feasible grid and HMMPs intersect):
              Generating candidate dummy trajectory point set Using Classical Algorithms
              If feasible grids and HMMPs and candidate dummy trajectory point set intersect:
                    Random selection of intersection point (feasible grids and HMMPs and
                    candidate dummy trajectory point set)
                    Generated = 1
              Else:
                          Random selection of intersection point (feasible grids and HMMPs)
                    Generated = 1
If not Generated:
        Random selection of point in feasible grids
        Adding the point to the dummy location dataset
Outputting  k − 1  dummy locations
Fusing 1 real-time location data with  k − 1  dummy locations
```

4.4 Experience and Analysis

Figure 8 shows the performance of each scheme against LSTM-based discriminator. Among them, MLN, MN and ADTGA do not consider the human mobility model, but all consider the background information. The experimental results show that the probability of our scheme being recognized is very low, that is, the generated dummy trajectory basically conforms to the human movement pattern.

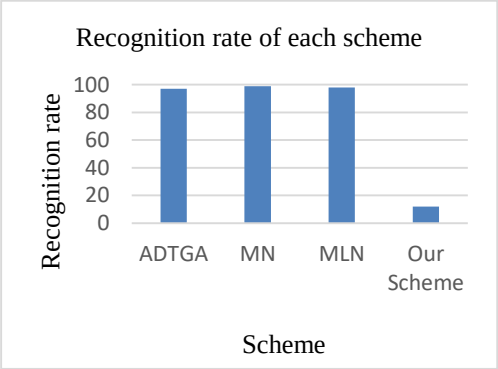


Fig. 8. Performance of each scheme against LSTM-based discriminator

4.5 Security Analysis

- ***Resistance to collusion attacks:*** collusion attacks occur when a group of users connect and share information. Our scheme does not need to cooperate with other parties, so it can resist collusion attacks.
- ***Resist to inference attacks:*** Inference Attack is an attack launched by an opponent based on prior knowledge such as map information and query probability. Our scheme takes full account of the map background information, and the selected trajectory points at each moment conform to the real trajectory's temporal and spatial distribution, so the scheme can resist inference attacks.
- ***Resist to channel attacks:*** Because the generation and publication of dummy trajectory points are in mobile intelligent terminals and do not pass through communication channels before the fusion of k trajectory point, the attacker can

  not obtain the real trajectory points from the channel end.

5 Summary

In this paper, we designed a LSTM-based dummy trajectory detection scheme and tested some trajectories generated by the classical dummy trajectory generation algorithm. The experimental results show that for the dummy trajectories generated by these algorithms, the average detection rate can reach more than 95%, and for real trajectories, the dummy positive rate is only about 10%. The experimental results show that the dummy trajectory generated by the classical algorithm does not conform to the human mobility model, that is, the spatio-temporal distribution that does not conform to the real trajectory. After that, we design a complete set of dummy trajectory generation schemes for the shortcomings of existing dummy trajectory generation algorithms. This scheme restricts the generated trajectory points to three major constraints. Because our design scheme can resist collusion attacks, inference attacks and channel attacks, it can effectively protect the user's trajectory privacy.

References

1. Krumm, John. (2009). A survey of computational location privacy. Personal and Ubiquitous Computing. 13. 391-399.10.1007/s00779-008-0212-5.
2. Wernke, M., Skvortsov, P., Dürr, F., & Rothermel, K. (2014). A classification of location privacy attacks and approaches. Personal and Ubiquitous Computing, 18(1), 163-175. doi:10.1007/s00779-012-0633-z
3. M. Tang, Q. Wu, G. Zhang, L. He and H. Zhang. A New Scheme of LBS Privacy Protection. 2009 5th International Conference on Wireless Communications, Networking and Mobile Computing, Beijing, 2009, pp. 1-6.

4. W. He. Research on LBS privacy protection technology in mobile social networks. 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, 2017, pp. 73-76.
5. Xu, Toby & Cai, Ying. (2009). Feeling-based location privacy protection for location-based services. Proceedings of the ACM Conference on Computer and Communications Security. 348-357. 10.1145/1653662.1653704.
6. T. Xu and Y. Cai. Exploring Historical Location Data for Anonymity Preservation in Location-Based Services. IEEE INFOCOM 2008 - The 27th Conference on Computer Communications, Phoenix, AZ, 2008, pp. 547-555.
7. Y. Wang, D. Xu, X. He, et al. L2P2: Location-aware location privacy protection for location-based services. 2012 Proceedings IEEE INFOCOM, Orlando, FL, 2012, pp. 1996-2004.
8. H. Kido, Y. Yanagisawa, and T. Satoh. An anonymous communication technique using dummies for location-based services. ICPS '05. Proceedings. International Conference on Pervasive Services, 2005, Santorini, Greece, 2005, pp. 88-97.
9. H. Kido, Y. Yanagisawa, and T. Satoh. Protection of Location Privacy using Dummies for Location-based Services. 21st International Conference on Data Engineering Workshops (ICDEW'05), Tokyo, Japan, 2005, pp. 1248-1248.
10. X. Wu, G. Sun. A Novel Dummy-Based Mechanism to Protect Privacy on Trajectories. 2014 IEEE International Conference on Data Mining Workshop, Shenzhen, 2014, pp. 1120-1125.
11. R. Kato, M. Iwata, T. Hara, et al. A dummy-based anonymization method based on user trajectory with pauses. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12). ACM, New York, NY, USA, 249-258.
12. B. Niu, Q. Li, X. Zhu, et al. Achieving k-anonymity in privacy-aware location-based services. IEEE INFOCOM 2014 - IEEE Conference on Computer Communications, Toronto, ON, 2014, pp. 754-762.
13. Jiaji Pan, Yining Liu, Weiming Zhang. Detection of dummy trajectories using convolutional neural networks. Security and Communication Networks. Vol. 2019, Article ID