International Journal of Information Security and Privacy
Volume 17 • Issue 1

Trajectory Data Publication Based on Differential Privacy

Zhen Gu, Harbin Engineering University, China*
https://orcid.org/0000-0003-1480-458X

Guoyin Zhang, Harbin Engineering University, China

ABSTRACT

Analyzing trajectory data can provide people with a higher quality of life. However, publishing trajectory data directly will leak privacy. The authors propose a trajectory data publication method based on differential privacy (TDDP). TDDP method consists of two stages. In the location generalization stage, firstly, the locations at each timestamp are clustered into classes by k-means++ algorithm, and then the representative location of each class is selected by using the exponential mechanism. In the generalized trajectory data publication stage, the authors design a sampling mechanism to form the generalized trajectories. The locations are sampled from the representative locations under different timestamps to form the generalized trajectories. The TDDP method can avoid the generation of non-semantic representative locations and ensure that the generalized trajectories can resist filtering attacks. The experimental results show that the trajectory data released by TDDP method can achieve a good balance between privacy protection and data availability.

KEYWORDS

Data Publication, Differential Privacy, Exponential Mechanism, Filtering Attacks, Generalized Trajectory, Privacy Leakage, Representative Location, Trajectory Data

INTRODUCTION

In the era of big data, location-aware technologies such as mobile communications and sensing devices digitize the geographic locations of people and objects, and subsequently generate a large amount of trajectory data. Location data contains characteristics of human behavior, by analyzing and mining trajectory data, better services can be provided to people (Yang et al., 2019). For example, urban traffic can be reasonably planned to avoid traffic congestion by analyzing trajectory data (Yuan et al., 2012, &Yuan et al., 2013). However, trajectory data contain a lot of sensitive personal information, such as the home address, work address, physical health status. If the location or trajectory data are directly released, it will lead to privacy leakage (Wernke et al., 2014, Gursoy et al., 2019, & Ding et al., 2020),

DOI: 10.4018/IJISP.315593

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

seriously, it will even threaten people's personal safety and property safety. The researches on trajectory data privacy protection are mainly divided into two types. One is the trajectory data privacy protection in offline mode. A specific organization collects trajectory data for analysis and mining to provide useful information to specific customers (Abul et al., 2008, Hua et al., 2015, Li et al., 2017, & Ma et al., 2021). The other type is online trajectory data privacy protection, such as location-based services. The real-time trajectory data of moving objects needs to be uploaded to the service provider, in this case, privacy protection of trajectory data is also required (Zhang et al., 2017, Zhang et al., 2018). In this paper, the authors mainly study the privacy protection of trajectory data in offline mode.

The existing trajectory data privacy protection methods mainly include k -anonymity method (Sweeney et al., 2002), encryption method and random disturbance method. The k -anonymity method is vulnerable to attacks with background knowledge. The encryption method is not a commonly used method due to its high computational cost. Among the random perturbation methods, the trajectory data publishing based on differential privacy has become a more popular research (Hua et al., 2015, & Liu et al., 2021).Differential privacy technology (Dwork et al., 2017) is the strongest unconditional privacy protection technology currently known, differential privacy can resist attacks from any background knowledge. However, some current researches on trajectory data publishing based on differential privacy also have some aspects that need to be improved.

(a) Some current researches require that the start and end times of any two trajectories must be the same or assume that the raw trajectories need to contain the same prefix or a common subsequence. However, it is difficult for the actual collected trajectory data to have these characteristics.
(b) Some current methods are to cluster the locations of all trajectories at each timestamp, and then use the cluster center of each class as a representative of all locations within the same class, at last, they use the cluster centers to generate the generalized trajectory. However, the cluster centers sometimes do not have semantic information, even, non-semantic representative locations can appear in multiple clusters, which will make the published trajectories to be identified and filtered by the adversary.

For the above aspects, the authors propose a trajectory data publishing method based on differential privacy (TDDP), and the contributions are as follows:

(a) The authors propose a trajectory data publishing method based on differential privacy. The TDDP method consists of two stages. In the first stage, the locations of each timestamp are clustered into K classes, and then the representative location of each class at each timestamp is selected by the exponential mechanism. In the second stage, locations are sampled from representative locations of different timestamps to form the generalized trajectories, which can avoid generating non-semantic locations and resist filtering attacks.
(b) In order to improve the utility of the trajectory data published, in the second stage, the TDDP method make the generalized trajectory data set is partly composed of the generalized trajectories which containing the raw trajectories.
(c) The authors conduct experiments on real data set. Hausdorff distance and spatial range queries are used for measuring the utility of the published trajectory data, mutual information is used to measure the total privacy loss, the results show that the published trajectory data can maintain good utility.

RELATED WORK

Abul et al. (2008) proposed the NWA (Never walk alone) algorithm which uses the inherent uncertainty of the motion trajectory to make the k trajectories in the trajectory cylinder indistinguishable and

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/trajectory-data-publication-based-on-differential-privacy/315593

Related Content

A Mutual Authentication Protocol with Resynchronisation Capability for Mobile Satellite Communications

Ioana Lasc, Reiner Dojenand Tom Coffey (2011). *International Journal of Information Security and Privacy (pp. 33-49).*

www.irma-international.org/article/mutual-authentication-protocol-resynchronisation-capability/53014

Computer Security Practices and Perceptions of the Next Generation of Corporate Computer Use

S.E. Kruckand Faye P. Teer (2008). *International Journal of Information Security and Privacy (pp. 80-90).*

www.irma-international.org/article/computer-security-practices-perceptions-next/2477

Competing on Performance on the Global Marketplace: Applying Business Analytics as a Robust Decision Tool

Rajagopal (2017). *Business Analytics and Cyber Security Management in Organizations (pp. 1-13).*

www.irma-international.org/chapter/competing-on-performance-on-the-global-marketplace/171831

The State-of-the-Art Cryptography Techniques for Secure Data Transmission

Bhanu Chander (2020). *Handbook of Research on Intrusion Detection Systems (pp. 284-305).*

www.irma-international.org/chapter/the-state-of-the-art-cryptography-techniques-for-secure-data-transmission/251807

Retrieval of Information Through Botnet Attacks: The Importance of Botnet Detection in the Modern Era

Zahian Ismail, Aman B. Jantan, Mohd. Najwadi Yusoffand Muhammad Ubale Kiru (2022). *Handbook of Research on Technical, Privacy, and Security Challenges in a Modern World (pp. 337-356).*

www.irma-international.org/chapter/retrieval-of-information-through-botnet-attacks/312430