

***Department of Computer Science & Engineering***

**Course Title –** Computer Networks

**Course Code –** CSE 319

**Section** - A

Final Report on Research

***Detailed Tittle B***

**Date of Submission –** May 13, 2022



## 

## **1. Introduction**

With the development of modern technology, many apps and software are using the user's location for their location-based services. These location-based services have generated a massive amount of mobility databases. Collecting consumer statistics has been a common practice of companies to understand consumers’ insights to improve services and products.

Crowdsourcing application [[2](#bookmark=id.17dp8vu)], GPS location and vehicle’s location data is an important asset to preserve security of the user. Nowadays almost every mobile application is capable of tracking users’ location by agreements. Crowdsourcing applications collect users’ location data by launching location-related tasks which are completed by the user to obtain rewards according to the application. For example, Pokémon GO (video games).

And for GPS tracking applications the authors [[3](#bookmark=id.3rdcrjn)] collect spatio-temporal information of the moving objects daily stored in the form of trajectories. Internet of connected vehicles communicate with each other which contain sensitive information of other users. This location-based service has generated a massive amount of mobility databases. While processing of location data also comes with threats on the privacy of the recorded users. This immense amount of data has the capability of producing patterns as the general population’s actions are heavily dependent on some common variables. Mining and analyzing these data may reveal individual personal information like real time location and personal information, which is risky for users.

To overcome these privacy issues, many efforts in literature aim to develop protection mechanisms. There are some frameworks that don't rely on third-party servers and users can automatically configure their privacy and utility. Data privacy is a term that implies the property of seclusion of sensitive information in a dataset. The aim of a privacy technique is to deter the identification of a person by anonymizing the individuals in the dataset to protect sensitive information, while the dataset can be used by the interested third parties for analysis purposes without obstruction.

## **2.Domain and Background**

The domain of these papers is mobility data privacy. These papers introduce solutions to problems related to this domain such as protecting privacy data about places of interest, daily schedule, based on previous data. Similarly, models have been developed to protect crowdsourced mobility information. Finally, privacy data protection requires there to be balance [[10](#bookmark=id.z337ya)] between the data lost in the privacy protection methods and the utility the manipulated data would provide.

## **3. Challenges**

Participants prefer to upload their trajectory data to the data collection server after they have been protected and processed, rather than through third-party servers for privacy protection [[11](#bookmark=id.3j2qqm3)]. So, if the data collection server is not trusted or attacked by an attacker, based on the historical trajectory data of the participants collected by the server, by analyzing and mining the mobility patterns of the participants, attackers can infer the sensitive information of the individual, such as home address and workplace, point of interest, living habits and even social attributes. The relevant research shows that the trajectory of moving objects often has a high temporal and spatial regularity. The most common threats on privacy [[7](#bookmark=id.1ksv4uv)] are **i)** re identification attacks where the identity of an anonymous user is guessed based on previously recorded data. **ii)** mobility prediction that anticipates the user's next moves based on their habits. **iii)** extraction of user’s places of interest (Home/ workplace) and **iv)** inference of social relationships. In that case adopting the same level of privacy-protection mechanism for all participants may result in unnecessary information and data loss. But the existing mechanisms focus on the trajectory of privacy protection of the mobility patterns.

Crowdsourcing platforms try to infer the users’ real-time actual position or even the future mobile behavior based on the observed locations, such as the locations of the LBS queries and the accepted. The spatiotemporal correlation hidden in the users’ mobile behaviors which can be derived from the historical mobile data. GPS and various location-based services collecting large scale spatio-temporal locations of moving objects by service providers, are stored in the form of trajectories. These trajectories contain users’ personal information like real time location. Mining and analyzing these trajectories an attacker may find the users real life pattern to harm their privacy. It may even result in serious threats to the safety of users’ life and property.

Local differential privacy [[9](#bookmark=id.2jxsxqh)] is a promising privacy preserving model for statistical aggregation of user data that prevents user privacy leakage from the data aggregator. However, such metrics cannot capture the question that which one is the optimal privacy mechanism in a set of equivalents -privacy mechanisms. Besides, the privacy and utility are closely correlated with the privacy mechanism, and existing methods do not consider the strategic adversary’s behavior. When noise is introduced into the data, it makes it harder for any malicious third party to identify the user from the data. However, it also pollutes the data and reduces the utility. Therefore, there is a compromise being made between data privacy and data utility value each time Differential privacy is applied.

## **4. Proposed Solutions**

The so-called LPPM only modifies the location information of users to improve their privacy level. For instance, GEO-I adds noise to the spatial information of a user data. Using LPPM [[5](#bookmark=id.lnxbz9)] may result in various levels of privacy and utility depending on the properties of the user's mobility. They proposed a framework PULP standing for Privacy and utility LPPM Parametrization. It automatically selects a LPPM among different ones, and determines the best configuration of the LPPM based on each user’s objective. PULP uses nonlinear models to capture the impact of each LPPM on data privacy and utility levels.

Again, a trajectory privacy-preserving framework based on a configurable participant terminal, named MSPP in mobile devices. MSPP doesn’t rely on third-party servers, and adaptively selects a corresponding privacy protection scheme based on individual historical trajectory data and privacy preferences stored locally by the participant terminal. Based on user target privacy requirements and data utility objects in different PS scenarios, MSPP provides participants with different TPPM.

They proposed a mobility-aware differential private solution: ConCrowdDP. Before starting application they constructed a spatiotemporal mobile model, STMarkov [[2](#bookmark=id.17dp8vu)], to model spatiotemporal correlation. Then, perturbed location is generated for the user to participate in the crowdsourcing application, according to STMarkov and *K*-norm DP. It is a mobility-aware differently private trace is generated for the user to participate in the application continually. For the trajectories, they are not simply considered as a sequence of the coordinates in Euclidean space, they combine the semantics-aware information with the background knowledge of underlying map for the location points. A cloaking region for each personalised sensitive place is built before a reasonable dummy trajectory in the cloaking region is constructed.

LDP [[8](#bookmark=id.44sinio)] achieves the privacy guarantee by introducing random noise into user data before transmitting them to the data aggregator while maintaining the users’ statistics to be accurate. Since the data aggregator cannot confidently know the raw user data, the users have plausible deniability and, therefore, their privacy remains protected to some degree. Currently, the LDP is widely investigated in the privacy protection area, and has been applied into privacy-preserving data collecting and analysing. The model takes in multiple parameters, one of which is called privacy loss or privacy budget, and takes a real positive value specified by the data aggregator to control the strength of the privacy protection. The smaller the value of the parameter, the higher the probability that the noisy representations of user values are the same or, in other words, the less likely the data aggregator realizes the true user values. According to the comparison of this paper the b-RAPPOR performed the best when privacy was lower, kRR when it was higher.

## **5. Future Scopes**

**A. Problem Proposal**

Most of the papers deal with the problem of local differential privacy by first mixing in noise in the current dataset. For example, the MSPP model changes the method for addition of noise based on the utility the certain person would provide. It would first observe the person’s location utility and then based on that it would introduce the noise to the data. Another example, differential privacy is used widely for privacy protection, which controls the noise using Laplace Mechanism. This also introduces noise to the system based on the data itself. This, however, poses the problem of being biased towards persons that fulfil certain criteria. This would result in a system that is not capable of providing a complete anonymous data solution.

To solve this problem, we can use a scheme that swap data between users, like Dynamic Pseudonym swap zone to protect location privacy of users. For each user, the scheme will allow the user to exchange the pseudonym with another random user in the just formed zone. Moreover, the scheme can self-adapt to the varying surroundings to reduce the communication cost in high density areas.

## **B. Solution Proposal**

To solve this problem, we can use a scheme that swap data between users, like Dynamic Pseudonym swap zone [[12]](#bookmark=id.1y810tw) to protect location privacy of users. For each user, the scheme will allow the user to exchange the pseudonym with another random user in the just formed zone. Moreover, the scheme can self-adapt to the varying surroundings to reduce the communication cost in high density areas. However, we can take this further by swapping information of a single movement among multiple users each. This, if executed with a proper optimization, may outperform the existing methods with better results.

## **6. Reference:**

[1] Yang, Q., Shen, Y., Vatsalan, D., Zhang, J., Kaafar, M. and Hu, W., 2020. P4Mobi: A Probabilistic Privacy-Preserving Framework for Publishing Mobility Datasets. IEEE Transactions on Vehicular Technology, 69(7), pp.6987-6999.

[2] Qiu, G. and Shen, Y., 2021. Mobility-Aware Differentially Private Trajectory for Privacy-Preserving Continual Crowdsourcing. IEEE Access, 9, pp.26362-26376.

[3] Ye, A., Zhang, Q., Diao, Y., Zhang, J., Deng, H. and Cheng, B., 2020. A Semantic-Based Approach for Privacy-Preserving in Trajectory Publishing. IEEE Access, 8, pp.184965-184975.

[4] Ghane, S., Jolfaei, A., Kulik, L., Ramamohanarao, K. and Puthal, D., 2021. Preserving Privacy in the Internet of Connected Vehicles. IEEE Transactions on Intelligent Transportation Systems, 22(8), pp.5018-5027.

[5] Cerf, S., Bouchenak, S., Robu, B., Marchand, N., Primault, V., Mokhtar, S., Boutet, A. and Chen, L., 2021. Automatic Privacy and Utility Preservation for Mobility Data: A Nonlinear Model-Based Approach. IEEE Transactions on Dependable and Secure Computing, 18(1), pp.269-282.

[6] Luo, Y., Jia, X., Fu, S. and Xu, M., 2019.Privacy-Preserving Ride Matching Over Road Networks for Online Ride-Hailing Service. IEEE Transactions on Information Forensics and Security, 14(7), pp.1791-1802.

[7] Xu, Z., Yang, W. and Wang, J., 2019. MSPP: A Trajectory Privacy-Preserving Framework for Participatory Sensing Based on Multi-Strategy. 2019 IEEE Globecom Workshops (GC Wkshps).

[8] Le, B. and Zia, T., 2021. Discrete Distribution Estimation with Local Differential Privacy: A Comparative Analysis. 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops),.

[9] Wu, N., Peng, C. and Niu, K., 2020. A Privacy-Preserving Game Model for Local Differential Privacy by Using Information-Theoretic Approach. IEEE Access, 8, pp.216741-216751.

[10] Esmeel, T., Hasan, M., Kabir, M. and Firdaus, A., 2020. Balancing Data Utility versus Information Loss in Data-Privacy Protection using k-Anonymity. 2020 IEEE 8th Conference on Systems, Process and Control (ICSPC),.

[11] D. Christin, ”Privacy in mobile participatory sensing: Current trends and future challenges”, J. Syst. Softw., vol. 116, pp. 57-68, June 2016.

[12] Yang, M., Feng, Y., Fu, X. and Qian, Q., 2019. Location privacy preserving scheme based on dynamic pseudonym swap zone for Internet of Vehicles. International Journal of Distributed Sensor Networks, 15(7), p.155014771986550.