Seeking Mr & Ms Regular: Sentinels to Characterize Crowd Dynamics (Vision Paper)

Elham Naghizade[1], Jeffrey Chan[2], and Martin Tomko[1][0000−0002−5736−4679]

[1] The University of Melbourne, VIC 3010, Australia
{enaghi, tomkom}@unimelb.edu.au
[2] RMIT University, VIC 3000, Australia
jeffrey.chan@rmit.edu.au

Abstract. Fine-grained GPS trajectory data are collected and mined at an unprecedented pace, revealing individuals' whereabouts, habits, and even personal preferences and values. They are used in personalized services as well as for analyses with societal benefits. Current privacy-preserving methods in the literature focus on protecting every single individual in the system. The large correlation between points in a trajectory and also between trajectory instances require, however, excessive levels of distortion to make the data private. Consequently, these methods either are too vulnerable to privacy attacks or cannot maintain data utility. To address this fundamental problem we are proposing a model that moves beyond *learning from as much data as possible* to a crowd-sourced, yet highly selective model that decides *who to learn from*. We believe that such a paradigm can specifically be applicable to detecting local anomalies, e.g., abnormal traffic congestion.

Keywords: Trajectory data · Individual movement profile · Traffic anomaly detection

1 Introduction

GPS-enabled navigation systems, smart phones and wearables are becoming omnipresent, ubiquitously sensing their users and capturing fine-grained mobility data [22, 23]. Mining the resulting large spatio-temporal datasets provides valuable insights for managing cities and the society in general [24]. In particular, fine-grained spatio-temporal data is indispensable for urban management and traffic monitoring [5, 7, 11].

Spatio-temporal data incorporate a large amount of personal, sensitive information, hence users are willing to control when, how and with whom their data is being shared. Several approaches have been proposed to safeguard user privacy when handling large-scale trajectory datasets [1, 6, 20]. Nonetheless, many of the proposed privacy-preserving approaches are either not applicable to real-world scenarios (e.g., encrypted trajectories), or are perturbing data excessively, thus limiting data utility [10, 12, 15, 25].

Traffic management is a domain where the public good of the data use is rarely contested. Yet, large-scale and highly privacy-invasive monitoring of users in the traffic is currently still necessary to assure valid analytical outcomes for fine-grained traffic monitoring tasks. While the ability to characterize coarse to mid-level, aggregate spatial and spatio-temporal dynamics of the urban traffic system has been extensively researched over the last decade [11, 2, 4, 16], here we focus on the nuanced, *local* understanding of the underlying population dynamics, since local traffic disturbances may result in major traffic congestions and timely detection of them is critical for traffic management.

To better preserve the privacy of the majority of users while detecting local traffic incidents, it may be possible to only observe *selected* users, thus preventing intrusive, fine-grained population-scale monitoring. A key question is then *whether a carefully selected group of data volunteers could act as a surrogate for the total population and thus enable capturing important local traffic patterns?*

We hypothesize that tracking a group of users with highly predictable movement patterns, denoted as *sentinels* in this paper, can provide sufficient information to finely characterize the anomalies and deviations in the traffic flow caused by e.g., an accident. These users can be identified and contacted in a targeted manner and encouraged to share their knowledge of these anomalies privately, possibly in exchange for remuneration or premium services. This approach significantly reduces the number of users that are required to be tracked, but also makes it possible to establish plausible incentives for data sharing.

Several research question will arise if this hypothesis is confirmed: What fraction of the population should be considered as sentinels? What should be the spatio-temporal distribution of sentinels to ensure valid traffic pattern detections? What type of local events can be revealed through observing the sentinels' behaviour? How to incentivize sentinels for the information exchange?

2 Related Work

Preserving trajectory privacy has gained a considerable attention in recent years [1, 6, 20]. Hand in hand, privacy implications of sharing large-scale trajectory data have become the focus of several studies. Starting from initial studies of the potential exposure of raw GPS trajectories and call-record datasets to privacy attacks [14, 9, 26, 17], more recent research shows the feasibility of attacking even seemingly *private* trajectory datasets. Gambs et al. [8] propose a de-anonymization attack inspired by the Netflix case [19]. They assume the adversary has access to the mobility patterns of the users for a short amount of time. Building a Mobility Markov Chain model, the authors then use this model to re-identify the users. In [13], the authors succeed at reconstructing an unknown private trajectory using its distance to a number of known trajectories, possibly with different sampling interval. The authors in [21] investigated the privacy implication of sharing the results of location-based queries like 'Where is the nearest gas station to my path?' or 'Send me the location of closest Italian restaurants.'. Hence, only a set of POIs is released which consists of the query

results at each timestamp. This set of POIs is then used to reconstruct the original trajectory of the user. The authors in [18] focused on personalized privacy guarantees and show that the coarse trajectory of private users can be refined using the fine-grained trajectory of privacy agnostic users. In all the studies above, the trajectories of all users are shared, thus increasing both the risk of privacy breaches and requiring a large amount of data distortion to make their trajectories private. In contrast, we propose a framework that uses a small sub-set of privacy-agnostic population to sense traffic anomalies.

3 The Vision of the Future System

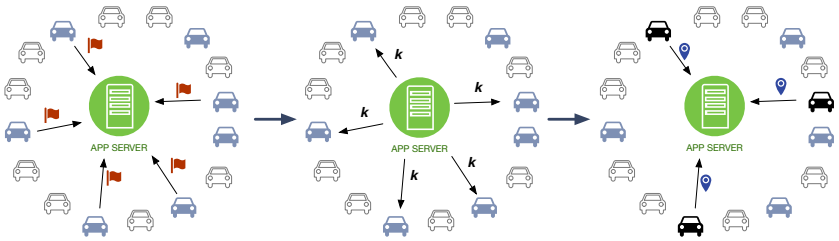


Fig. 1: The overall view of our proposed framework. White cars represent the general population of cars, light gray cars represent sentinels, black cars represent sentinels that have responded to the incentive and shared their information.

Our framework focuses on the detection of outlying spatio-temporal events, (e.g., an unusual traffic congestion pattern). Our privacy-aware traffic detection system has two key features: First, it is built on the knowledge that is shared by a small subset of the population, so-called *sentinels*. The sentinels have highly predictable routes and are either privacy-agnostic users, or less private agents, e.g., delivery drivers, with flexibility in planning their routes in case of a traffic incident. In our system, sentinels are willing to share limited location information with the server, e.g., through incentives such as special services. Moreover, our system allows the adaptation of typical k-anonymous privacy-preserving measures. The benefits of achieving k-anonymity in such scenario is two-fold: i) safeguarding the privacy of sentinels, and ii) reaching a confidence level to detect meaningful events and reduce the noise.

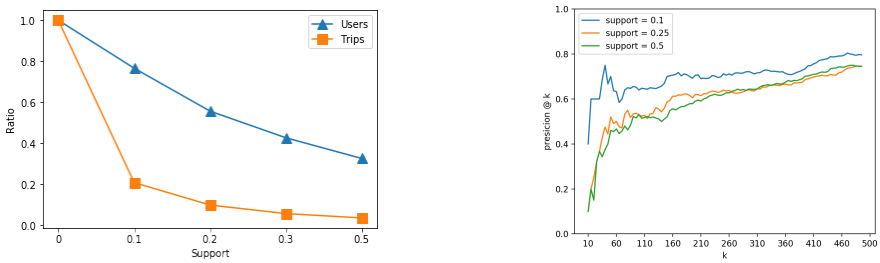
We furthermore envisage light-weight algorithms operating on sentinels' personal devices to build their movement profile and submit the time and location of a deviation from their profile. Figure 1 provides an overview of our system. Each user is assigned a confidence score for the current route based on the predictability of that route given his/her movement history. Note that this score is computed locally on users' device and the server does not have access to the fine-grained trajectory of users. Given the scores, users are divided into sentinels

and the rest of the general traffic population (Figure 1). In the event of a deviation from the movement profile, the server is notified (Figure 1, left). The server then responds with the potential incentive as well as the required confidence and privacy level, i.e., k (Figure 1, middle). The system then receives the time and location of the local deviation from k sentinels (Figure 1, right).

4 Preliminary Experimental Feasibility Study

We examine the feasibility of the envisaged system through a preliminary experimental study and explore the following questions:

1. ***Are there users that can become our sentinels?*** We examined a 10-month GPS trajectory dataset in the city of Melbourne. We selected users with at least 50 trips, which cover distances larger than $3km$ and are longer than 15 minutes (2707 users). Focusing only on their fully sampled trips, we have 31748 trips. Figure 2a shows the ratio of users and their trips that have sufficient support, where support is defined as the number of trips with the same origin and destination divided by the total number of user's trips.



(a) (b)

Fig. 2: (a) Effect of support on the ratio of sentinels. (b) Precison@k.

2. ***What is the spatial coverage of their trips?*** We divided the space into grids ($\approx 1 \times 1km$) and counted the number of GPS points that falls into each grid cell. Setting support to 0.1 and 0.5 leads to 72% and 60% of cell coverage respectively. Figure 2b shows the precision @ k [3] when querying about the top k populated grids. As can be seen, using sentinels' information provides promising levels of understanding of the general traffic dynamics.

5 Conclusion

We propose a novel privacy-aware system that can help detect local traffic anomalies using a small sub-set of population, i.e., sentinels, instead of every user's data. The underlying idea of our system is that a traffic anomaly can cause sentinels to deviate from their normal movement pattern. We explored the feasibility of such approach by investigating whether or not sentinels exist in real-world datasets and if they are reliable for estimating the traffic condition.

Bibliography

[1] Abul, O., Bonchi, F., Nanni, M.: Never walk alone: Uncertainty for anonymity in moving objects databases. In: 2008 IEEE 24th International Conference on Data Engineering. pp. 376–385 (2008)

[2] Ahas, R., Silm, S., Jrv, O., Saluveer, E., Tiru, M.: Using mobile positioning data to model locations meaningful to users of mobile phones. Journal of Urban Technology 17(1), 3–27 (2010)

[3] Buckley, C., Voorhees, E.M.: Evaluating evaluation measure stability. SIGIR Forum 51(2), 235–242 (Aug 2017), http://doi.acm.org.ezp.lib.unimelb.edu.au/10.1145/3130348.3130373

[4] Calabrese, F., Di Lorenzo, G., Liu, L., Ratti, C.: Estimating origin-destination flows using opportunistically collected mobile phone location data from one million users in boston metropolitan area. IEEE Pervasive Computing 10(4), 36–44 (2011)

[5] Castro, P.S., Zhang, D., Li, S.: Urban traffic modelling and prediction using large scale taxi gps traces. In: Kay, J., Lukowicz, P., Tokuda, H., Olivier, P., Krüger, A. (eds.) Pervasive Computing. pp. 57–72 (2012)

[6] Damiani, M.L., Bertino, E., Silvestri, C.: Protecting location privacy against spatial inferences: The probe approach. In: Proceedings of the 2Nd SIGSPATIAL ACM GIS 2009 International Workshop on Security and Privacy in GIS and LBS. pp. 32–41. SPRINGL '09, ACM, New York, NY, USA (2009)

[7] D'Andrea, E., Marcelloni, F.: Detection of traffic congestion and incidents from gps trace analysis. Expert Systems with Applications 73, 43 – 56 (2017), http://www.sciencedirect.com/science/article/pii/S0957417416306935

[8] Gambs, S., Killijian, M.O., del Prado Cortez, M.N.: De-anonymization attack on geolocated data. Journal of Computer and System Sciences 80(8), 1597 – 1614 (2014), special Issue on Theory and Applications in Parallel and Distributed Computing Systems

[9] Gruteser, M., Hoh, B.: Security in Pervasive Computing: Second International Conference, SPC 2005, Boppard, Germany, April 6-8, 2005. Proceedings, chap. On the Anonymity of Periodic Location Samples, pp. 179–192. Springer Berlin Heidelberg, Berlin, Heidelberg (2005)

[10] He, X., Cormode, G., Machanavajjhala, A., Procopiuc, C.M., Srivastava, D.: Dpt: Differentially private trajectory synthesis using hierarchical reference systems. Proc. VLDB Endow. 8(11), 1154–1165 (2015)

[11] Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q., Bayen, A.M.: Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. Transportation Research Part C: Emerging Technologies 18(4), 568 – 583 (2010), http://www.sciencedirect.com/science/article/pii/S0968090X09001430

[12] Jiang, K., Shao, D., Bressan, S., Kister, T., Tan, K.L.: Publishing trajectories with differential privacy guarantees. In: Proceedings of the 25th In-

ternational Conference on Scientific and Statistical Database Management. pp. 12:1–12:12. SSDBM, ACM, New York, NY, USA (2013)
[13] Kaplan, E., Pedersen, T.B., Savas, E., Saygin, Y.: Discovering private trajectories using background information. Data and Knowledge Engineering 69(7), 723 – 736 (2010)
[14] Krumm, J.: Inference attacks on location tracks. In: PerComp, pp. 127–143 (2007)
[15] Liu, A., Zhengy, K., Liz, L., Liu, G., Zhao, L., Zhou, X.: Efficient secure similarity computation on encrypted trajectory data. In: 2015 IEEE 31st International Conference on Data Engineering. pp. 66–77 (April 2015)
[16] Lovelace, R., Ballas, D., Watson, M.: A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. Journal of Transport Geography 34, 282 – 296 (2014), http://www.sciencedirect.com/science/article/pii/S0966692313001361
[17] de Montjoye, Y.A., Hidalgo, C.A., Verleysen, M., Blondel, V.D.: Unique in the Crowd: The privacy bounds of human mobility. Scientific Reports 3, 1376 (Mar 2013)
[18] Naghizade, E., Bailey, J., Kulik, L., Tanin, E.: How private can i be among public users? In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 1137–1141. UbiComp '15, ACM, New York, NY, USA (2015), http://doi.acm.org.ezp.lib.unimelb.edu.au/10.1145/2750858.2805836
[19] Narayanan, A., Shmatikov, V.: How to break anonymity of the netflix prize dataset. CoRR abs/cs/0610105 (2006), http://arxiv.org/abs/cs/0610105
[20] Nergiz, M.E., Atzori, M., Saygin, Y.: Towards trajectory anonymization: a generalization-based approach. In: Proceedings of the ACM SIGSPATIAL 2008 International Workshop on Security and Privacy in GIS and LBS. pp. 52 – 61 (2008)
[21] Quattrone, A., Naghizade, E., Kulik, L., Tanin, E.: Tell me what you want and i will tell others where you have been. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. pp. 1783–1786. CIKM '14, ACM, New York, NY, USA (2014)
[22] Ratti, C., Frenchman, D., Pulselli, R.M., Williams, S.: Mobile landscapes: using location data from cell phones for urban analysis. Environment and Planning B: Planning and Design 33(5), 727–748 (2006)
[23] Ren, Y., Tomko, M., Salim, F., Ong, K., Sanderson, M.: Analyzing web behavior in indoor retail spaces. Journal of the American Society for Information Science 68(1), 62–76 (2017)
[24] Ren, Y., Tomko, M., Salim, F.D., Chan, J., Sanderson, M.: Understanding the predictability of user demographics from cyber-physical-social behaviours in indoor retail spaces. EPJ Data Science 7(1), 1–21 (2018)
[25] Wernke, M., Durr, F., Rothermel, K.: Pshare: Position sharing for location privacy based on multi-secret sharing. In: PerCom. pp. 153–161 (2012)
[26] Zang, H., Bolot, J.: Anonymization of location data does not work: A large-scale measurement study. In: MobiCom. pp. 145–156 (2011)