Attaquer/protéger des données de localisations (agrégées).

Jean-François Couchot

1 Présentation

Depuis une décennie environ, les données relatives à la mobilité humaine sont collectées continûment par les réseaux cellulaires, les applications mobiles, les montres de sport et sont souvent ensuite publiées à des fins de recherche ou commerciaux. On pense aux cartes de chaleur comme celles d'*Uber movement*[1], aux données publiées par FluxVision chez Orange[2]

Comme la publication de telles données sur la mobilité des individus est en conflit avec la confidentialité, les dépositaires de ces ensembles de données ont souvent tendance à ne publier que des données de mobilité agrégées. Par exemple, peut être publié le nombre d'utilisateurs sur une zone d'intérêt vaste heure par heure. Ceci peut être considéré comme suffisant pour préserver la confidentialité des utilisateurs.

Il semble que cette famille de données puisse être attaquée [TXL+18]. Les auteurs de ce travail ont montré que par une extension de l'algorithme d'affectation hongroise, on pouvait reconstruire des trajectoires uniques d'utilisateurs avec une grande probabilité. Votre trajectoire est unique, reconstructible, même si vous ne le pensiez pas !

2 Proposition de travail

Dans un premier temps, il s'agit de comprendre les données qui sont au coeur de ce problème. Le groupe travaillera sur

— le jeu de données d'*Uber movement* et ce qu'ont peut en faire [PR21], par exemple ;
— les données FluxVision et ce qu'ont peut en faire [PVS+20], par exemple
— de trouver d'autres cartes de chaleur. . . et de les exploiter.

Dans un second temps, il s'agira d'implanter une attaque du type [TXL+18] (mais ce n'est pas la seule) sur le jeu de données d'Uber par exemple. Un gros travail sur les données prises en entrée devra être réalisé, par exemple sur le mesocentre de calcul de l'UFC.

Dans un troisième temps, il s'agira de proposer une méthode permettant de mettre en défaut ce genre d'attaque. On pourra par exemple penser à ajouter du bruit aux données pour rendre impossible, sinon très difficile l'affectation hongroise.

Un stage de recherche financé pourra être proposé à l'issu de ce travail.

Références

[PR21] Jane Perlman and Shouraseni Sen Roy. Analysis of human movement in the miami metropolitan area utilizing uber movement data. *Cities*, page 103376, 2021.

[PVS+20] Giulia Pullano, Eugenio Valdano, Nicola Scarpa, Stefania Rubrichi, and Vittoria Colizza. Population mobility reductions during covid-19 epidemic in france under lockdown. *MedRxiv*, 29 :2020, 2020.

[TXL+18] Zhen Tu, Fengli Xu, Yong Li, Pengyu Zhang, and Depeng Jin. A new privacy breach : User trajectory recovery from aggregated mobility data. *IEEE/ACM Trans. Netw.*, 26(3) :1446–1459, 2018.

1. `https://movement.uber.com/explore/san_francisco/mobility-heatmap/query`
2. `https://www.orange-business.com/en/products/flux-vision`