

Anonymisation : le cas MTL Trajet

L'utilisation de technologies récentes en transport ouvre des possibilités opérationnelles, mais permet également de recueillir des informations détaillées sur les utilisateurs des réseaux routiers. Ces données sont de plus en plus rendus disponibles publiquement ("données ouvertes") ou partagées avec des chercheurs et autres partenaires. Elles doivent alors être anonymisées pour assurer la confidentialité des usagers.

Cependant, les techniques utilisées dans le domaine des transports sont souvent simples et intuitives mais sont inadéquates pour protéger l'anonymat. En effet, un article de 2013 de MontJoye¹ montre qu'à partir d'un ensemble de données de téléphonie cellulaire, il suffit de 4 points d'une traces spatio-temporelle pour identifier 95% des usagers, c'est-à-dire caractériser de façon unique chacun des usagers dont les données ont été collectées.

Cet article présentera les résultats d'une étude sur l'anonymat des données publiques de l'ensemble recueillie avec l'application MTL Trajet. MTL Trajet est une application mobile, lancée par la Ville de Montréal en 2016, qui enregistre les déplacements, peu importe le mode de transport utilisé. En 2017, plus de 4000 personnes avaient généré 21 millions de points GPS sur l'ensemble du réseau. L'objectif de recueillir ces données est de permettre une optimisation opérationnelle de la ville.

La méthode employée par la Ville pour anonymiser les traces GPS des usagers est de tronquer le début et la fin des trajets afin d'éliminer la possibilité de retrouver le domicile ou encore le lieu de travail des usagers. L'extrémité de chaque trajet (origine et destination) est tronquée à l'intersection empruntée la plus proche. Par contre, cette méthode ne comporte pas de notion de distance minimale de troncage entre l'intersection et les débuts et fins de trajet, ni, surtout, de densité de trajet commençant ou se terminant à une intersection à un moment donné.

Ainsi, cet article présentera les résultats de désanonymisation des données publiques de MTL Trajet collectées en 2017 suivant Montjoye et. Al. De plus, l'utilisation d'autres sources de données (par ex. Twitter) pourrait permettre d'identifier des usagers. À noter qu'ici la présentation des résultats, eux, ne permettront pas d'identifier les usagers. Finalement, l'article présentera la méthode et les résultats de la K-anonymité qui permet de garantir qu'un individu dans un ensemble de données ne puisse être identifié à partir des k-1 individus qui sont également inclus dans l'agrégation dans laquelle ils se trouvent.

¹ de Montjoye, Y.-A.; Hidalgo, C. A.; Verleysen, M. & Blondel, V. D. (2013), 'Unique in the Crowd: The privacy bounds of human mobility', *Scientific Reports* **3** .