

Privamov : Suivi d'utilisateurs mobiles et reconstruction de trajectoires

Patrice Raveneau ^{*} and Andrea Caballero ^{*} and Razvan Stanica ^{*} and Hervé Rivano ^{*}

^{*} *Université de Lyon, INRIA, INSA-Lyon, CITI-INRIA, F-69621, Villeurbanne, France*[†]

Afin de mieux penser les réseaux du futur, il est nécessaire de comprendre la relation qu'ont les utilisateurs avec le réseau, ainsi que la mobilité urbaine. Cette analyse peut être réalisée grâce à l'étude de données collectées au travers de divers capteurs du téléphone (GPS, modules sans fil, etc.). Une utilisation intensive de ces éléments diminue l'autonomie des téléphones mobiles. Cet article étudie l'impact de la diminution de la fréquence d'utilisation de la puce GPS sur la précision de la reconstruction de la trajectoire des utilisateurs. Nous avons distribué des téléphones munis d'une application permettant de collecter ces informations en continu. Suite à ces déploiements, nous pouvons simuler une fréquence d'activation du GPS moins élevée en sous-échantillonnant les données de localisation collectées. Nous avons ensuite eu recours à des méthodes d'interpolation linéaire et cubique pour reconstruire la trajectoire des utilisateurs. Nous faisons varier les valeurs des périodes de sous-échantillonnage et analysons comment l'erreur d'interpolation évolue avec l'augmentation de la période de sous-échantillonnage. Nous réalisons une première étude sur la totalité de la mesure, puis nous restreignons aux périodes où l'utilisateur n'est pas statique. Les résultats obtenus montrent que la collecte d'une mesure par quart d'heure permet de reconstruire la trajectoire avec un niveau d'erreurs acceptable.

Keywords: GPS, Estimation de trajectoires, Crowd-sensing, Expérience, Mesures

1 Introduction

Depuis plusieurs années, les téléphones mobiles sont en mesure de déterminer leur position. Le marché des services reposant sur la localisation représentera d'ici 2016 près de 10 milliards de dollars [ZLJG13]. Posséder cette information de localisation peut s'avérer pratique pour l'utilisateur en mesure de trouver ce dont il a besoin à proximité du lieu où il se trouve, ainsi que pour les annonceurs, pouvant proposer des produits et services susceptibles d'intéresser un utilisateur.

Des problèmes de respect de la vie privée peuvent apparaître avec la transmission de ces données de localisation [dMHVB13]. Les utilisateurs de téléphones mobiles n'ont pas forcément conscience que leur téléphone est capable de transmettre en permanence leur position vers différents serveurs et ceci créant un risque potentiel pour leur vie privée. De plus, un utilisateur conscient de ces risques n'a actuellement pas une pléthore d'options pour s'en prémunir. Soit, il fait confiance aux instances qui ont accès à ses données, soit il doit se passer des avantages d'un téléphone mobile, s'il souhaite maîtriser un peu plus les données transmises à son sujet.

Concernant les capteurs de localisation du téléphone, l'utilisation du GPS est gourmande en énergie [BBKN15]. D'autres moyens peuvent être envisagés pour réaliser une localisation à moindre coût, mais moins précise, avec l'empreinte cellulaire ou des réseaux sans fil vue du téléphone [PKSG11]. Le recours à un système collaboratif, reposant sur la correspondance entre les empreintes de connectivité et les localisations GPS transmises par de nombreux utilisateurs et sans conservation d'identifiants, pourrait fournir un service de localisation moins consommateur en énergie et respectant plus la vie privée, si l'utilisateur a la main sur la transmission des données.

[†]Ce travail a été réalisé grâce au soutien financier du LABEX IMU (ANR-10-LABX-0088) de l'Université de Lyon, dans le cadre du programme "Investissements d'Avenir" (ANR-11-IDEX-0007) géré par l'Agence Nationale de la Recherche (ANR).

Cependant, certaines applications, comme les applications de navigation, nécessitent un certain suivi pour pouvoir guider un utilisateur vers sa destination. Si l'utilisateur s'écarte du chemin proposé, il faut être en mesure de le détecter pour lui proposer un itinéraire adapté à sa nouvelle trajectoire. Afin d'estimer si un système de localisation collaboratif serait efficace, nous étudions le lien entre période d'échantillonnage de la position et erreur de reconstruction de trajectoire. Nous serons ainsi en mesure de fournir une mesure de la qualité de la localisation et d'indiquer à partir de quelle valeur une localisation n'est plus assez précise pour un certain type de service.

Nous présentons dans un premier temps l'architecture de la plate-forme de collecte et son paramétrage, puis nous utilisons nos données GPS brutes pour créer des jeux de données sous-échantillonnées sur la totalité de la collecte dans un premier temps, puis en se limitant aux périodes où l'utilisateur est mobile. Une fois ce sous-échantillonnage effectué, nous interpolons les points restants et comparons les trajectoires obtenues aux trajectoires initiales afin d'évaluer l'évolution de l'erreur avec ce paramètre.

2 Collecte de données

L'objectif du projet Privamov[‡] est de fournir une plate-forme de collecte de données d'utilisateurs mobiles afin d'analyser différentes caractéristiques propres aux réseaux de téléphonie mobile. La mobilité des utilisateurs, leur activité réseau en fonction des heures de la journée ou du lieu où ils se situent, leurs habitudes de transport ; ces différentes informations peuvent être obtenues à partir de l'analyse des données recueillies par les capteurs d'un téléphone moderne.

Ces données ne peuvent être conservées indéfiniment sur un téléphone sans quoi il finirait par manquer de mémoire ; ni ne peuvent être transmises directement vers un serveur si le téléphone n'est pas raccordé à un réseau quelconque. Nous avons donc fait le choix de développer une application Android qui réalise un compromis entre ces deux modes de fonctionnement. Les données captées sont tout d'abord stockées sur la mémoire du téléphone, puis sont rapatriées périodiquement vers un serveur dès lors que le téléphone est raccordé à un réseau WiFi. Nous avons déjà réalisé un premier déploiement en décembre 2014, lors de la conférence ACM/IFIP/USENIX Middleware[§], qui se tenait à Bordeaux. Les données collectées lors de ce déploiement font partie des traces analysées dans cette étude. Avant le déploiement à la conférence Middleware, nous possédions des traces en provenance de 10 utilisateurs correspondant à plus de 400 heures de suivi. Après ce déploiement, nous avons désormais 25 utilisateurs et la durée totale des traces analysées représente plus de 35 jours.

Pour cette étude, nous nous limitons à l'analyse des données en provenance de la puce GPS. Nous maintenons le GPS activé en permanence et, pour chaque téléphone s , nous nous concentrons sur l'analyse du triplet position : $\mathbf{p}_s = (\text{longitude}, \text{latitude}, \text{date})$

3 Étude de la diminution de la fréquence d'activation du GPS

Le but de notre étude est d'estimer le niveau d'erreur introduit par la diminution de la période d'activation du GPS. Afin d'estimer cette erreur, nous simulons, pour chaque téléphone s , d'autres traces dont la période entre deux positions successives est supérieure à un seuil arbitraire. Nous notons D_s , l'ensemble des dates appartenant à une trace T_s et $\mathbf{p}_s(t)$ la position à l'instant t . Nous appelons l'ensemble des positions d'un téléphone s une trace, T_s , et nous nommons une trace sous-échantillonnée, $T_s(d)$, où d est la période minimale en secondes entre deux positions successives ; $D_s(d)$ est l'ensemble des dates appartenant à $T_s(d)$. Nous avons donc : $\forall (d > 1) ; \text{card}(T_s(d)) \leq \text{card}(T_s)$

3.1 Étude avec tous types de mobilités

Nous faisons varier la période d'échantillonnage de 10 à 900 secondes. Une fois le sous-échantillonnage réalisé, nous utilisons une interpolation linéaire ou cubique pour reconstruire les positions supprimées lors du sous-échantillonnage. Nous rappelons les expressions des polynômes d'interpolation $P(x)$:

Linéaire	$P(x)_{x \in [x_i, x_{i+1}]} = a + b \cdot x$	calculé à partir des $(x_j, y_j), j \in [i ; i + 1]$
Cubique	$P(x)_{x \in [x_i, x_{i+1}]} = a + b \cdot x + c \cdot x^2 + d \cdot x^3$	calculé à partir des $(x_j, y_j), j \in [i ; i + 3]$

‡. <http://liris.cnrs.fr/privamov/>

§. <http://2014.middleware-conference.org/>

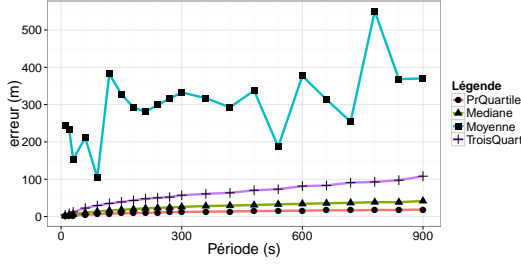


FIGURE 1: Statistiques de l'erreur d'interpolation cubique

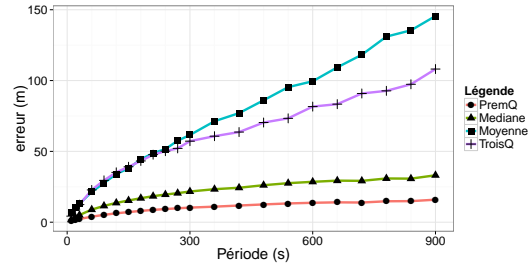


FIGURE 2: Statistiques de l'erreur d'interpolation linéaire

L'ensemble des positions estimées, $\widehat{T}_s(d)$, nommé trace interpolée du sous-échantillonnage d , est obtenu de la sorte, où $Interpol(\mathbf{p}_s(t))$ est la position interpolée à l'instant t :

$$\widehat{T}_s(d) : \{\mathbf{p}_s(t)\}_{t \in \{D_s(d)\}} \cup \{Interpol(\mathbf{p}_s(t))\}_{t \in \{D_s \setminus D_s(d)\}}$$

Nous avons donc un ensemble original de positions et un ensemble de positions estimées de même taille. De plus, $\{\widehat{D}_s\} = \{D_s\}$. Nous comparons ensuite les traces obtenues avec la trace originale et mesurons l'erreur pour chaque position de la trace interpolée. Le calcul de l'erreur est effectué comme suit pour chaque position de la trace, où $dist(\mathbf{x}, \mathbf{y})$ est la distance euclidienne entre \mathbf{x} et \mathbf{y} :

$$err(t) = dist(\mathbf{p}_s(t), \widehat{\mathbf{p}}_s(t)) ; \forall t \in \{D_s\}$$

Les figures 1 et 2 présentent ces résultats.

Le premier résultat que l'on observe est que, pour les deux types d'interpolation, la moyenne est supérieure au troisième quartile. On en déduit donc que la majorité du temps, les erreurs d'interpolation sont faibles, mais que quelques points ont une erreur très grande et ces erreurs font augmenter la moyenne.

Un autre résultat notable est que les erreurs restent relativement faibles pour toutes les périodes d'échantillonnage observées. En effet, pour l'interpolation linéaire, la moyenne reste en-dessous de 160 mètres même pour une période de 15 minutes ; dans le cas de l'interpolation cubique, la moyenne ne dépasse pas les 400 mètres.

On constate également que l'interpolation linéaire, bien que plus simple, fournit de meilleurs résultats que l'interpolation cubique. L'interpolation cubique présente d'ailleurs un comportement plus erratique que l'interpolation linéaire dont les erreurs semblent augmenter linéairement avec la période d'échantillonnage.

3.2 Étude avec mobilité filtrée

Ces résultats présentant des erreurs très faibles et connaissant les déplacements des personnes suivies, nous avons été amenés à penser que les longues périodes d'inactivité pouvaient créer un biais dans nos résultats. C'est pourquoi, nous avons décidé de filtrer les périodes où la mobilité est nulle à faible, pour ne conserver des traces qu'avec des mobilités moyennes à élevées.

Ayant constaté que l'interpolation linéaire fournissait de meilleurs résultats que la cubique, nous limitons notre étude à cette interpolation. La figure 3 présente les résultats obtenus de la fonction de répartition de l'erreur pour les différentes périodes d'échantillonnage sélectionnées.

On constate ici que les résultats restent très bons pour les faibles périodes de sous-échantillonnage. En effet, lorsque la période d'échantillonnage est réglée à deux minutes, 90% des erreurs sont inférieures à cent mètres. Une précision de l'ordre de la centaine de mètres est plus qu'acceptable pour la majorité des applications ayant besoin de l'information de localisation.

En revanche, les erreurs ne sont pas aussi insensibles à l'augmentation de la période que dans la première étude. Lorsque l'on atteint les 15 minutes pour la période d'échantillonnage, on retrouve moins de 80% des erreurs inférieures à 200 mètres.

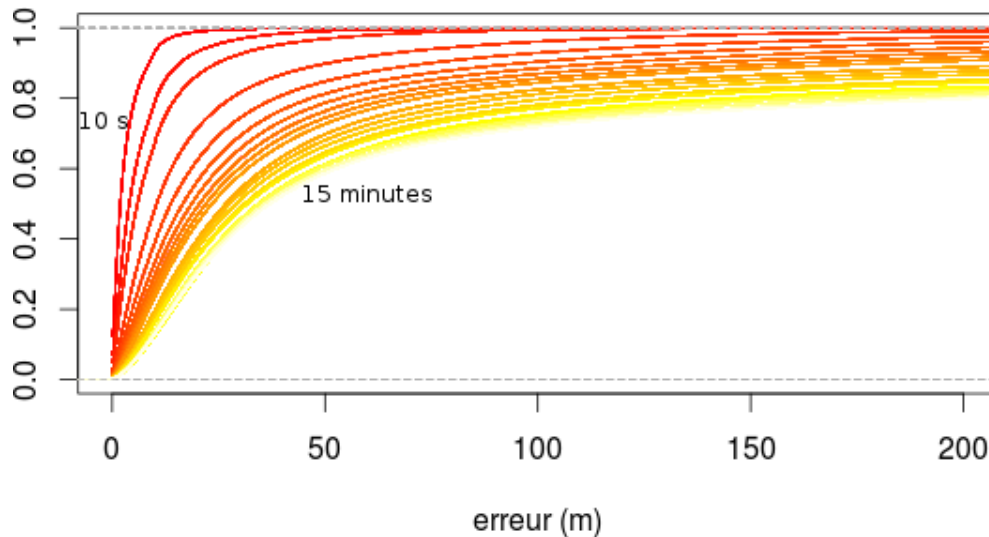


FIGURE 3: Évolution de la fonction de répartition de l'erreur d'interpolation pour différentes périodes d'échantillonnage (de 10s, en haut à gauche, à 15 minutes, en bas à droite)

À partir des résultats de nos déploiements, il semblerait qu'une période de 60 secondes garantirait un bon ratio de reconstruction de trajectoires avec des pertes minimales sur la localisation. De plus, nous obtenons de très bons résultats sur la reconstruction de trajectoires avec de longues périodes d'inactivité du GPS. L'ajout d'autres informations, comme la correspondance avec les routes d'une carte, pourrait augmenter la précision de la reconstruction.

4 Conclusion

En conclusion, les données collectées par notre plate-forme nous ont permis de constater qu'il était toujours possible de reconstruire presque entièrement une trajectoire avec très peu d'erreurs pour des périodes d'inactivation du GPS pouvant aller jusqu'à plusieurs minutes.

Nous avons également constaté que des mécanismes d'interpolation simples permettent d'obtenir un bon niveau de reconstruction de trajectoires. De plus, nous pensons pouvoir affirmer, que pour des études de ce type, il est nécessaire d'identifier et de supprimer les périodes d'inactivité de l'analyse, le côté sédentaire des utilisateurs induisant un biais dans les résultats.

Il semblerait néanmoins que sur les périodes où l'utilisateur reste sédentaire, l'extinction du GPS peut être augmentée. Un tel fonctionnement permettrait d'économiser de la batterie, sans perdre d'information. Une activation/désactivation du GPS reposant sur d'autres capteurs, tels que l'accéléromètre ou l'environnement radio, semble une bonne perspective.

Références

- [BBKN15] S. Bhattacharya, H. Blunck, M.B. Kjærgaard, and P. Nurmi. Robust and energy-efficient trajectory tracking for mobile devices. *Mobile Computing, IEEE Transactions on*, 14(2) :430–443, 2015.
- [dMHVB13] Yves-Alexandre de Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the crowd : The privacy bounds of human mobility. *Scientific reports*, 3, 2013.
- [PKSG11] Jeongyeup Paek, Kyu-Han Kim, Jatinder P. Singh, and Ramesh Govindan. Energy-efficient positioning for smartphones using cell-id sequence matching. In *Proceedings of ACM MobiSys 2011*, pages 293–306, 2011.
- [ZLJG13] Lei Zhang, Jiangchuan Liu, Hongbo Jiang, and Yong Guan. Senstrack : Energy-efficient location tracking with smartphone sensors. *Sensors Journal, IEEE*, 13(10) :3775–3784, 2013.