Campagne de collecte de données et vie privée

Nicolas Haderer, Miguel Nuñez del Prado Cortez, Romain Rouvoy,
Marc-Olivier Killijian, Matthieu Roy

Campagne de collecte de données et vie privée

Nicolas Haderer[1], Miguel Núñez, del Prado Cortez[2,3], Romain Rouvoy[1], Marc-Olivier Killijian[2], and Matthieu Roy[2,3]

[1] INRIA Lille – Nord Europe, Project-team ADAM
University Lille 1, LIFL – CNRS UMR 8022, France
{nicolas.haderer,romain.rouvoy}@inria.fr
[2] LAAS-CNRS, France
[3] Université de Toulouse ; UPS, INSA, INP, ISAE ; LAAS ; F-31077 Toulouse, France
{mnunezde,mkilliji,mroy}@laas.fr

Résumé Les communautés scientifiques ont souvent recours à la simulation dans le but de valider leurs théories. Cependant, la pertinence des résultats obtenus est fortement dépendante de la qualité des traces générées par les simulateurs. Ce phénomène est particulièrement vrai lorsque l'on considère les traces de mobilité humaine qui sont difficilement prévisibles. Dans ce contexte, la popularité des nouvelles générations de smartphones, équipés d'une grande variété de capteurs (GPS, bluetooth, accéléromètre, etc.), offre de nouvelles perspectives pour la collecte de données réalistes au sein d'une population. Cependant, la nature sensible, du point de vue de la vie privée, des informations collectées représente un des principaux obstacles aux déploiement généralisé d'une application de collecte de données et à son adoption auprès des utilisateurs.

C'est pourquoi nous présentons UBILAB , une nouvelle plate-forme permettant aux scientifiques de mettre en place facilement des campagnes de collecte de données et d'inférer automatiquement différentes attaques sur les données partagées par les utilisateurs mobiles afin de les avertir d'un risque potentiel d'atteinte à leurs informations privées.

1 Introduction

La nouvelle génération de smartphones (Android, iPhone), maintenant équipée d'une grande variété de capteurs (GPS, bluetooth, accéléromètre, etc.), offre de nouvelles perspectives à diverses communautés scientifiques afin de réaliser différentes campagnes de collectes de données massives d'une population et de son environnement. Ces données peuvent ainsi être exploitées pour mieux comprendre les mouvements d'une population, de mettre au point de nouveaux protocoles de communication, d'analyser les interactions sociales des utilisateurs, etc. La nature sensible des données collectées, généralement couplant des informations temporelles et géographiques, peuvent révéler des informations critiques sur la vie privée d'un utilisateur (résidence privée, opinion politique ou réseau social), même si celles-ci ont été préalablement anonymisées. Ce risque potentiel représente un des principaux obstacles aux déploiement généralisé d'une application de collecte de données et à son adoption auprès des utilisateurs.

Dans ce contexte, nous présentons UBILAB , une plate-forme dédié à la gestion de campagnes de collecte de données auprès d'utilisateurs de téléphones mobiles. UBILAB est le résultat de la l'association de deux plate-formes : ANTDROID [2] et GEPETO (GeoPrivacy Enhanced TOolkit)[1]. Ce système profite ainsi de l'architecture de ANTDROID pour rapidement mettre en place une campagne de collecte de données, et des algorithmes de GEPETO pour inférer automatiquement différentes attaques sur les données partagées par les utilisateurs mobiles afin de les avertir d'un risque potentiel d'atteinte à leurs vies privées.

2 UBILAB

La plate-forme est composée de deux parties — chacune est destinée aux différents acteurs évoluant dans la plate-forme : les scientifiques et les cobayes. Le serveur d'application dédié, destiné aux scientifiques, repose sur le style architectural REST (*REpresentational State Transfer*) fournissant l'ensemble des services pour la définition, la diffusion et l'exploitation d'une expérience de collecte de données. L'application cliente est une application Android téléchargeable, utilisée par une communauté d'utilisateurs pour partager leurs traces d'activités. Pour ce faire, l'utilisateur s'abonne à une ou plusieurs campagnes publiées par des scientifiques. Ces campagnes correspondent à des scripts de collecte automatique des informations requises par le scientifique. Ces scripts sont téléchargés via le serveur d'application dédié puis interprétés par un moteur de script intégré dans l'application cliente. Afin de maîtriser toute diffusion d'information, l'application cliente dispose de différents contrôles permettant aux utilisateurs d'autoriser ou non la collecte de certaines informations jugées trop sensibles (par ex., sa position).

Les données collectées par le téléphone mobile peuvent ensuite être envoyées manuellement par l'utilisateur ou automatiquement lorsque le téléphone est alimenté en courant pour limiter sa consommation énergétique. Avant d'être disponible pour les scientifiques, les données sont stockées dans une base de données temporaire ou un ensemble d'analyses sera effectué par la plate-forme GEPETO (GeoPrivacy Enhanced TOolkit) permettant de détecter si les données partagées peuvent comporter des risques vis-à-vis de la vie privée de l'utilisateur. Le résultat de ces analyses est ensuite renvoyée à l'utilisateur afin qu'il puisse évaluer si les données collectées peuvent compromettre sa vie privée. L'utilisateur peut alors ensuite décider de valider les données pour les rendre directement disponible pour les scientifiques, de les supprimer, ou d'appliquer des algorithmes d'assainissement (distorsion aléatoire, sous échantillonnage, etc..) fourni par GEPETO pour ajouter du bruit sur les traces de mobilité.

Références

1. S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez. Gepeto : a geoprivacy-enhancing toolkit. *AINA'09 Workshop on Advances in Mobile Computing and Applications : Security, Privacy and Trust, Perth, Australia.*, August 2009.
2. N. Haderer, R. Rouvoy, and L. Seinturier. AntDroid : A distributed platform for mobile sensing. Rapport de recherche RR-7885, INRIA, Lille, France., February 2012.