

L'ANONYMAT DIGITAL N'EXISTE PAS, AUTANT LE SAVOIR UNE BONNE FOIS POUR TOUTES

Publié le 24 juillet 2019



par Christian Du Brulle

Toutes les traces que nous générons, toutes les données nous concernant, que nous communiquons d'une manière ou d'une autre, même les plus anonymes et les plus incomplètes, toutes ces informations permettent... de nous identifier avec une quasi-certitude! Telle est la conclusion d'[une recherche menée par Luc Rocher, aspirant FNRS au pôle en ingénierie mathématique de l'UCLouvain](#).

Le doctorant a travaillé avec un spécialiste du domaine, le [Pr Yves-Alexandre de Montjoye, désormais professeur assistant, responsable du Computational Privacy Group à l'Imperial College](#) de Londres.

Les deux chercheurs, accompagnés par Julien Hendrickx ([lcteam/UCLouvain](#)) ont élaboré un

algorithme qui permet d'estimer, avec grande précision, si des données anonymes, mais « réidentifiées », appartiennent bien à une même personne ou non. L'algorithme évalue la probabilité pour une combinaison de caractéristiques connues d'être suffisamment précise pour décrire un seul individu parmi plusieurs milliards de personnes.

Une réidentification à 99,98%

En utilisant cette méthode, les chercheurs de l'UCLouvain et de l'Imperial College London ont montré que 99,98% des Américains seraient correctement réidentifiés dans n'importe quelle base de données en utilisant 15 attributs démographiques, avec des chiffres similaires à travers le monde (16 attributs en ajoutant la nationalité).

« Beaucoup de personnes vivant à New York sont des hommes et ont la trentaine. Parmi elles, beaucoup moins sont également nées le 5 janvier, conduisent une voiture de sport rouge, ont deux enfants et un chien », précise Luc Rocher, dans un communiqué de l'UCLouvain. « Des informations plutôt standards, que les entreprises demandent régulièrement. Et qui permettent de réidentifier les individus ».

Après avoir appris quelles caractéristiques rendent les individus uniques, les algorithmes des chercheurs génèrent des populations synthétiques pour estimer si un individu peut se démarquer parmi des milliards de personnes.

« En Belgique, depuis 2017, certaines données médicales collectées par des hôpitaux sont revendues de manière anonyme. Le sont-elles vraiment ? C'est là tout l'intérêt du modèle développé par les chercheurs de l'UCLouvain et de l'Imperial College London, puisqu'il permet désormais de vérifier la réalité, ou non, de cet anonymat », estime l'UCLouvain.

Tout le monde recueille des données

Voici un peu plus d'un an, l'ingénieur civil Yves-Alexandre de Montjoye, qui jongle avec les mathématiques appliquées et la protection de la vie privée avait déjà indiqué à Bruxelles, lors d'une conférence « [Science & Cocktails](#) », qu'au départ de quelques bribes de données, il était possible d'identifier avec quasi certitude un individu.

À quoi peuvent servir les données d'un téléphone portable ? Comment utiliser en toute sécurité les données volumineuses tout en allant de l'avant ? Ces questions étaient au centre de cette rencontre avec un large public.

« Nous vivons à une époque où l'information sur la plupart de nos mouvements et de nos actions est recueillie et stockée en temps réel. Tout le monde recueille des données sur vous : vos recherches sur Google, vos commandes de nourriture en ligne, vos lieux de vacances et les profils d'autres personnes que vous consultez sur Facebook. Et la disponibilité d'un téléphone mobile à grande échelle, d'une carte de crédit, d'un historique de navigation, etc., augmente considérablement notre capacité à comprendre et potentiellement affecter le comportement des individus et des collectifs.

Toutefois, l'utilisation de ces données soulève des préoccupations légitimes en matière de protection de la vie privée. Lors de cet événement, Yves-Alexandre de Montjoye a expliqué

comment les mécanismes traditionnels de protection des données ne parviennent pas à protéger la vie privée des personnes à l'ère des grandes données. Et il explique comment des informations sensibles peuvent souvent être déduites de données apparemment inoffensives.

Les mots de passe et la cryptographie n'y changent rien

Et pour celles et ceux qui pensent qu'avec quelques bons mots de passe et autres outils de cryptographie, il est possible de protéger durablement ses données... Mieux vaut tout de suite faire son deuil de cette idée.

« Qu'il s'agisse de messages, de photos de vacances ou de tout autre type d'informations personnelles transmises ou stockées sur internet: ces données sont absolument transparentes. Ou du moins, elles le seront dans un proche avenir», martèle le scientifique québécois Gilles Brassard, spécialiste de la cryptographie quantique.

De passage à Bruxelles en début d'année dans le cadre de la chaire du Québec à l'Académie royale des Sciences, ce professeur au Département d'informatique et de recherche opérationnelle de l'Université de Montréal disait: [«rien de ce que vous confiez en ligne n'est sûr. Ou plus exactement, rien de ce que vous avez confié ou transmis jusqu'à présent, et ce depuis les débuts de l'internet, n'est confidentiel](#) »...