

La solution dans une décentralisation extrême ?

Si les pratiques de la NSA ont causé un scandale international, les agences de renseignement françaises ne sont pas en reste pour la surveillance téléphonique et électronique de grande envergure. Moins d'un mois après les révélations d'Edward Snowden, l'interception des communications privées en France, aux frontières de la légalité, était dénoncée par les médias.

Cette dérive résulte de la centralisation des données, comme l'explique Alex Pentland dans cet article, et du fait qu'il soit devenu trivial de les consulter et de les croiser. Au-delà des pratiques des agences de renseignement, l'ère des données massives est confrontée à des enjeux contradictoires : comment à la fois publier des données d'utilité publique (recensement, fréquentation du métro, géolocalisation pour le covoiturage, etc.) et protéger la vie privée des individus ?

Depuis près d'une dizaine d'années, l'équipe Systèmes d'information sécurisés et mobiles d'Inria (Institut national de recherche en informatique et

en automatique), à Paris, dont je suis membre associé, travaille sur le projet PlugDB (pour *Plug Database*, littéralement base de données à brancher). Ce projet se fonde sur des principes proches de ceux présentés par A. Pentland, en particulier la décentralisation des données, tout en allant plus loin sur la sécurité et la confiance. L'élément central du dispositif est un petit appareil sécurisé connectable à un ordinateur. Il permet aux individus de stocker leurs données ou des liens sur ces données, et surtout d'en contrôler l'accès, le partage, la modification et la destruction.

C'est au sein même de ce dispositif hautement sécurisé

que s'exécutent les applications de gestion de données, et non dans un serveur d'Internet ou sur un ordinateur. Il faudra alors que les concepteurs d'applications fournissent les logiciels nécessaires pour les faire fonctionner. Le dispositif est sécurisé même contre l'utilisateur : quand ce dernier souhaite utiliser un logiciel, il est assuré que c'est bien lui qui s'exécute, mais il ne peut ni voir ni modifier le code sous-jacent. Les concepteurs sont alors libres de publier ou non le code, pour des raisons de transparence ou de secret industriel.

Prenons l'exemple de l'application *Google Flu Trends*, qui permet de prévoir les épidémies de grippe. Même si Google garantit contractuellement l'anonymat des utilisateurs et explique n'avoir besoin que de données agrégées (par exemple le nombre de fois que certains mots clés apparaissent dans une zone géographique),

il stocke en réalité de manière centralisée l'intégralité des requêtes posées, ainsi que différents paramètres liés à l'utilisateur (telle l'adresse IP). Et ce pendant plusieurs années ! Ces données sont à la merci d'un piratage et d'un usage abusif par Google ou par les États faisant pression sur l'entreprise.

Avec un instrument comme PlugDB, l'utilisateur n'a pas besoin de dévoiler ses données personnelles, dont il reste maître. Il peut choisir la précision de celles qui sont divulguées, surveiller les accès et leur fixer une date limite, etc. La confiance est permise par l'utilisation d'un système d'exploitation transparent (dont le code est librement consultable) et de matériel sécurisé, qui n'autorise pas la falsification des données, même par leur détenteur.

— Benjamin Nguyen
Institut national des sciences
appliquées (INSA), Bourges

où des milliers de milliards de dollars transitent chaque jour. Grâce notamment à ses systèmes de contrôle de métadonnées et de vérification automatique, ce réseau de confiance repousse les cambrioleurs et assure un trajet fiable des fonds jusqu'à la destination souhaitée.

Autrefois, les réseaux de confiance étaient complexes et leur fonctionnement coûteux, mais la démocratisation des capacités de calcul les a rendus accessibles à de petites organisations, voire à des individus. En partenariat avec l'Institut de conception guidée par les données, à Boston, mon équipe de l'Institut de technologie du Massachusetts (MIT) a participé à la construction de centres de données personnelles ouverts, ou *open PDS* (pour *Personal Data Store*), une version grand public de ce type de système. À l'heure actuelle, nous les testons avec des partenaires industriels et gouvernementaux. L'idée est de permettre à tous un partage de données sensibles, notamment sanitaires et financières, avec un niveau de

sécurité équivalent à celui de SWIFT. Quand les réseaux de confiance se répandront, il deviendra plus sûr d'échanger des données, dont l'enregistrement dans des architectures de stockage réparties et sécurisées protégera à la fois les individus et les organisations ; ainsi, les mauvaises utilisations des données massives devraient être évitées.

PRINCIPE 3 Toujours expérimenter

Le dernier principe, peut-être le plus important, est d'admettre que nous n'avons pas toutes les réponses et qu'il n'y a pas de solution définitive. Nous devons continuellement nous adapter. L'ère du numérique est totalement nouvelle, de sorte que nous ne pouvons pas nous contenter de méthodes classiques.

Sous la pression des autres pays, des citoyens et des entreprises du Web, la Maison Blanche a déjà proposé de limiter

la surveillance exercée par la NSA. Dans un effort pour restaurer la confiance, les entreprises du Web cherchent à obtenir le droit de diffuser des informations sur les requêtes de l'agence – des métadonnées sur des métadonnées. En mai 2014, la Chambre des Représentants a voté une loi qui, bien que jugée insuffisante par de nombreux défenseurs de la vie privée, commençait à restreindre les collectes massives de données et à introduire un peu plus de transparence (cette loi a cependant été refusée par le Sénat en novembre).

Ces mesures vont dans la bonne direction, mais nous ne pourrions trouver que des solutions à court terme. Les techniques évoluent et les procédures gouvernementales doivent suivre le rythme. Finalement, le changement le plus important serait que nous décidions de conduire en continu des expériences et des tests à petite échelle, afin d'innover sans cesse et de trier ce qui fonctionne de ce qui ne fonctionne pas. ■