Outils informatiques pour la protection de la vie privée

Guillaume Piolle *

24 juin 2011

En 2006, des chercheurs travaillant sur le moteur de recherche d'AOL publièrent (à fins d'expérimentation) trois mois d'historique de requêtes, en remplaçant les identifiants des utilisateurs par des numéros pour protéger leurs clients. Cette « anonymisation » insuffisante permit néanmoins d'identifier assez facilement un grand nombre de personnes, en exposant explicitement leurs centres d'intérêts, parfois très sensibles, par le biais de leurs recherches web. Le public se rendit compte à cette occasion, de manière assez soudaine et désagréable, de la quantité d'information qu'il confiait à certains acteurs d'Internet et du risque que cela pouvait représenter pour sa vie privée. Même en Europe, où les législations héritières de la loi « Informatique et Libertés » avaient instauré une certaine culture de la protection des données personnelles, Internet tel qu'il avait évolué représentait un nouvel espace de risques pour la vie privée. Aujourd'hui, les utilisateurs dans leur ensemble paraissent de plus en plus concernés par cette problématique, convaincus que la protection de la vie privée sur Internet est un aspect essentiel de leur liberté individuelle et non simplement un moyen de cacher des actions honteuses ou répréhensibles. Cela peut s'exprimer de manière très dramatique dans la nécessité, pour les dissidents d'un régime autoritaire, de protéger leur anonymat en ligne.

Cependant, pour qui cherche sa place dans la nouvelle ère numérique, le respect de la sphère privée est un enjeu omniprésent, mais délicat à appréhender. Peuvent en témoigner par exemple les efforts significatifs demandés en 2009 à Facebook par les autorités canadiennes pour permettre aux utilisateurs de mieux protéger leur profil. Si l'on peut généralement élaborer ses propres règles de fonctionnement quant à son exposition sur Internet, il est parfois difficile d'anticiper les risques liés à l'outil ou au réseau eux-mêmes. Heureusement, il existe de nombreux outils,

*Supélec, équipe SSIR, CS 47601, Avenue de la Boulaie, 35576 Cesson-Sévigné Cedex, France – `guillaume.piolle@supelec.fr`

utilisables par tout un chacun avec plus ou moins de facilité, permettant de se prémunir contre un certains nombre de ces risques (par exemple en masquant son adresse IP, en gérant mieux ses mots de passe ou en chiffrant sa connexion, ses données ou ses messages). Ce sont ces outils auxquels nous nous intéressons ici.

1 Le monde numérique, source de risques pour la vie privée

Les risques relatifs aux données personnelles ne sont pas, dans leur grande majorité, directement dûs à l'informatique. Ils résultent davantage des décisions politiques et organisationnelles prises par les administrations et les entreprises, et le but premier de la loi Informatique et Libertés est l'encadrement de ces décisions. Cependant, l'outil informatique et Internet en particulier favorisent grandement la collecte de données à l'insu de l'utilisateur et l'interconnexion des sources de données. La densité des activités humaines dans le monde numérique constitue donc un terrain fertile pour de potentielles atteintes à la vie privée des utilisateurs.

Les risques techniques, qui nous placent dans le contexte plus général de la sécurité informatique, sont de diverses natures. Il existe tout d'abord la possibilité, pour un certain nombre d'acteurs sur Internet (fournisseurs de services, d'accès, prestataires techniques...), de tracer les activités d'un utilisateur donné. Cette attention particulière portée aux communications de l'utilisateur peut s'effectuer à différents niveaux techniques et avoir des objectifs de nature et de légitimité très différents, allant des obligations légales des fournisseurs d'accès à l'envoi de publicités ciblées, en passant par les enquêtes judiciaires et le profilage comportemental. Il existe bien évidemment un risque que des communications électroniques soient interceptées, espionnées ou modifiées avec des intentions délibérément malveillantes. Ce risque est généralement considéré comme minime par les particuliers, néanmoins les outils permettant d'automatiser ces interceptions sont malheureusement de plus en plus faciles à utiliser. Enfin, il existe toujours un risque de vol physique des données, sous la forme d'un ordinateur ou d'un support externe. Ce dernier risque concerne tous les types d'utilisateurs.

À la plupart de ces risques techniques correspondent des outils plus ou moins efficaces, qu'un utilisateur peut mettre en œuvre directement pour protéger sa vie privée et ses données personnelles.

2 Protéger et anonymiser sa connexion sur Internet

Le premier aspect concerne la protection de la connexion informatique qui s'instaure entre le poste de l'utilisateur et un serveur distant lorsqu'un site web est visité, qu'un courrier électronique est envoyé ou lors de toute autre activité sur le réseau. Différentes techniques permettent d'assurer différentes propriétés sur cette connexion, comme la protection du contenu des messages échangés ou

l'anonymisation de l'adresse IP de l'utilisateur (adresse qui permet de déduire de nouvelles informations, notamment relatives à la géolocalisation de l'utilisateur).

La technologie SSL/TLS, en jeu pour l'accès aux URL de type `https://`, établit un canal de communication chiffré entre le navigateur de l'utilisateur et un serveur (typiquement un serveur web ou de courrier électronique). L'avantage de cette technologie est d'être intégrée à tous les navigateurs web et de ne nécessiter aucune action de l'utilisateur (sinon la lecture d'éventuels avertissements). SSL/TLS met en place ce que l'on appelle un chiffrement point-à-point : les deux extrémités ont accès au contenu en clair, mais les intermédiaires transmettant les messages (ainsi que toute personne capable d'écouter la connexion) n'en voient que la version chiffrée, a priori incompréhensible. On assure donc la **confidentialité** de la communication. Néanmoins, l'existence de la connexion entre les deux machines (identifiables par leur adresse IP) reste visible. De plus, on ne se protège pas de son interlocuteur, qui a nécessairement accès au contenu en clair. Pendant la révolution du jasmin en janvier 2011, les services de renseignement tunisiens furent fortement soupçonnés d'avoir obtenu frauduleusement les identifiants de connexion des opposants qui n'utilisaient pas SSL/TLS pour se connecter aux services de Google et Facebook.

Il existe également des moyens de modifier l'adresse IP de l'utilisateur, pour contribuer à l'anonymisation de son activité. Il est possible de passer par un serveur mandataire, ou *proxy*, qui servira d'intermédiaire pour l'ensemble du trafic web ou d'une application donnée. Suivant sa configuration, le serveur mandataire peut présenter sa propre IP en lieu et place de celle de l'utilisateur. Plus efficace, on peut avoir recours à un réseau privé virtuel, ou VPN (pour *Virtual Private Network*). C'est un service (souvent payant) permettant à l'utilisateur de se connecter à une machine donnée qui servira d'intermédiaire, tout en chiffrant les communications entre l'utilisateur et le fournisseur de VPN. Dans ce cas, il n'est pas possible pour un observateur extérieur d'associer une activité donnée sur le réseau à l'adresse IP de l'utilisateur. Néanmoins, il faut avoir confiance dans le fournisseur du service, qui a accès à l'ensemble des communications en clair.

L'utilisation du réseau Tor, elle, fait passer la communication par un certain nombre d'intermédiaires, choisis aléatoirement, en utilisant un chiffrement « en couches » assez complexe (Tor signifie *The Onion Router* en référence à ces couches de chiffrement) avant de la transmettre à son destinataire. Dans ce système, personne ne peut a priori associer le contenu du message à l'adresse IP de l'utilisateur, et ce dernier peut empêcher tout rapprochement entre deux activités différentes sur Internet. Néanmoins, l'utilisation de Tor nécessite d'installer et de configurer le logiciel adéquat et ralentit considérablement la navigation.

3 Gérer ses identités en ligne

Un aspect tout aussi essentiel de la vie privée sur Internet concerne la gestion des identités. Cela est d'autant plus important que nombre d'activités en ligne

(comme l'utilisation de forums ou de plates-formes collaboratives) laissent une trace plus ou moins pérenne et publiquement accessible.

Le plus souvent, l'utilisateur se sert d'un pseudonyme. Il peut d'ailleurs en avoir plusieurs, qui servent à des familles d'utilisation voulues distinctes et cloisonnées. Cependant, il est très courant que le même pseudonyme serve sur plusieurs sites web différents. Deux risques surviennent alors : la mise en péril de la sécurité des mots de passe et le rapprochement entre les différentes activités menées sous un même pseudonyme.

En effet, si l'utilisateur se sert du même identifiant sur plusieurs sites, le risque est grand d'utiliser également le même mot de passe. Ce serait un choix désastreux : si l'un des administrateurs de l'un des sites n'est pas digne de confiance, il pourrait compromettre ce mot de passe et obtenir un accès aux les autres sites web, voire au compte de courrier électronique de l'utilisateur. D'un autre côté, il est très difficile de maintenir un ensemble de mots de passe différents et sûrs pour l'ensemble des sites web utilisés. Les solutions de type *Single-Sign-On* permettent de ne s'authentifier qu'une seule fois auprès d'un unique service, pour avoir ensuite accès à tout un ensemble de ressources. Un seul mot de passe est donc nécessaire. OpenID est le protocole de ce type le plus utilisé sur le web. Concrètement, l'utilisateur ouvre un compte (associé à un mot de passe) auprès d'un fournisseur de service OpenID de son choix, qui met à sa disposition une URL qui sera son identifiant. Lorsque l'utilisateur visite un site compatible, il s'identifie en fournissant cette URL. Le site web, le navigateur et le fournisseur OpenID interagissent alors automatiquement pour authentifier l'utilisateur. Le site web visité n'utilise donc tout simplement pas de mot de passe.

Si cette solution améliore la vie privée de l'utilisateur en évitant les risques liés aux mots de passe multiples, elle n'empêche pas les différents sites web de partager leurs informations de connexion et de relier entre elles les différentes activités d'un même utilisateur, ce qui peut également être une atteinte à sa vie privée. Certaines solutions techniques encore expérimentales, comme la technologie U-prove de Microsoft, devraient permettre d'empêcher cela dans une certaine mesure, mais elles ne sont pas encore aisément utilisable.

4 Assurer la confidentialité et l'intégrité des données

Le chiffrement et la signature cryptographiques sont deux outils de sécurité informatique exploitées comme briques constitutives d'autres technologies, mais pouvant également être utilisées telles quelles par les utilisateurs. Si ces outils sont très liés, leurs objectifs sont distincts. Le chiffrement permet d'assurer la **confidentialité** des données, en faisant en sorte que les personnes non destinataires ne puissent accéder au contenu en clair. De son côté, la signature en assure **l'intégrité** : la vérification de la signature garantit à la fois l'identité du signataire et le fait que le message n'a pas été modifié depuis sa signature.

La technique du chiffrement peut évidemment être utilisée en dehors d'Inter-

net, sur un disque dur ou tout autre support. De nombreux logiciels existent, pour tous les systèmes d'exploitation. Il convient de s'assurer que l'algorithme utilisé est réputé sûr et que sa mise en œuvre technique est exempte de failles. Ces vérifications ne sont pas forcément à la portée de l'utilisateur, mais certains logiciels ont d'ores et déjà acquis une bonne réputation. C'est le cas par exemple de TrueCrypt (gratuit et disponible sur tous les systèmes d'exploitation courants), devenu célèbre en 2010 lorsque le FBI a échoué à décrypter un disque dur chiffré par ce moyen.

Chiffrement et signature sont en outre largement mis en oeuvre dans le cadre des courriers électroniques. Les utilisateurs disposent de deux grandes familles d'outils pour chiffrer et signer leurs e-mails. Certains employeurs ou organisations fournissent parfois à leurs employés des certificats personnels, attachés à leur adresse électronique professionnelle, et dont l'authenticité est garantie par une chaîne de signatures remontant à une « autorité de confiance ». Ces certificats, que savent gérer la plupart des clients de messagerie modernes, peuvent servir à signer et parfois à chiffrer les messages. Une méthode plus souple consiste à utiliser la technologie PGP/GPG, qui nécessite en général l'utilisation d'un module complémentaire, comme Enigmail sur le logiciel Thunderbird. Ce système permet à l'utilisateur de générer et de contrôler ses propres clés de chiffrement et de signature, en les attachant à n'importe quelle adresse e-mail (potentiellement anonyme ou pseudonyme). L'authenticité des signatures est alors garantie par un système de réseau de confiance entre utilisateurs, chacun pouvant signer la clé d'un autre pour attester de son authenticité.

5 Contrôler l'utilisation des données personnelles

Jusqu'à présent, les techniques présentées avaient principalement pour but d'éviter la collecte indue de données personnelles ou d'autres informations confidentielles. Cependant, il est rare que l'on arrive à éviter totalement la fuite de données, d'autant que l'on transmet souvent soi-même volontairement certaines informations. On peut alors vouloir contrôler l'usage qui est fait de ces données, afin d'en interdire une exploitation trop intrusive.

De manière générale, il est malheureusement très difficile, voire impossible, de contrôler parfaitement l'utilisation qui est faite des données une fois qu'elles ont été confiées à des tiers. Cet aspect constitue actuellement un enjeu majeur pour la recherche dans le domaine de la protection de la vie privée. Il reste possible, néanmoins, d'influer sur certains aspects symptomatiques.

Une configuration appropriée du navigateur web est sans doute un premier moyen. On peut tout d'abord lutter partiellement contre le profilage effectué par les régies publicitaires en ligne, en interdisant les *cookies* déposés par les sites tiers (ceux déposés par le site visité sont généralement inévitables si l'on veut profiter pleinement du service). De plus, il convient d'examiner les réglages de conservation d'historique de son navigateur (URL visitées, cache, mots de passe, remplissage

de formulaires) et de s'assurer qu'ils conviennent à nos besoins particuliers et au risque perçu (notamment, pour un portable, en cas de perte ou de vol). Il n'y a malheureusement pas de recette toute faite : chacun doit déterminer ses propres réglages. La protection des mots de passe enregistrés par un mot de passe « maître » est toutefois un passage obligé. On peut également s'intéresser avec bénéfice aux modes de navigation privée et de nettoyage de l'historique fournis par les navigateurs modernes. Ils permettent, respectivement, de naviguer sans rien mémoriser en local et de supprimer tout ou partie des données conservées. Enfin, la publicité en ligne, ciblée ou non, peut facilement être filtrée par certains modules complémentaires des navigateurs. Le logiciel AdBlock est probablement parmi eux le plus efficace et le plus simple d'utilisation.

De manière similaire, les clients de messagerie modernes, ainsi que les fournisseurs d'accès et de *webmails* dans une moindre mesure, disposent de fonctionnalités permettant de filtrer les courriers indésirables. Après une nécessaire période d'apprentissage, pendant laquelle le logiciel prend acte de ce que vous considérez ou pas comme un message légitime, les courriers indésirables pourront être mis en quarantaine, supprimés à vue ou étiquetés, à la convenance de l'utilisateur.

6 Pour une « hygiène numérique » des données personnelles

Tous les outils et techniques que nous venons de présenter peuvent servir à contrer des menaces de nature principalement technique, informatique. Certains sont ouvertement dédiés à faire échouer des tentatives d'intrusion ou de vol de données délibérées et probablement mal intentionnées. Néanmoins, la part des données personnelles divulguée à cause de raisons techniques ou d'actions malveillantes est généralement négligeable par rapport à la quantité d'information volontairement transmise par l'utilisateur. Il est en effet aisé de chercher à se protéger d'un risque de vol de données fantasmé sans se rendre compte que les informations de profil que nous publions volontairement sur tel ou tel site sont également sensibles. Ce comportement est typiquement accentué par le phénomène psychologique qui nous fait exagérer un bénéfice attendu à court terme et sous-estimer un risque potentiel sur le long terme.

En conclusion, la maîtrise des techniques présentées ici n'est donc qu'un complément utile d'une « hygiène numérique » des données personnelles, qui doit conduire un utilisateur à ne divulguer, de manière raisonnée, que les informations nécessaires à la réalisation d'un objectif donné. Cela nécessite de développer une certaine conscience de la valeur que peuvent avoir des données apparemment anodines et donc du risque qu'elles représentent.