

“Sur Internet, s’est super, on peut surfer comme on veut, on est complètement anonyme. Personne ne sait qui est qui.” Cette phrase maintes fois entendue ne reflète qu'une partie de la réalité. Sur le réseau, tous les faits et gestes des internautes sont enregistrés et peuvent être étudiés et ensuite utilisés à des fins plus ou moins avouables. Nous allons ici présenter quelques techniques qui, utilisées par des personnes mal intentionnées peuvent être dévastatrices pour la sphère privée des internautes.

Le principal outil utilisé pour récolter des données est le *cookie*. Il s'agit d'une petite information qui est envoyée depuis le serveur vers le browser pour y être stockée. Elle sera ensuite incluse automatiquement dans toute requête du browser vers ce même serveur (et vers aucun autre). Ce principe peut être utilisé par le serveur pour stocker un peu d'information sur la personne qui surf, par exemple sa langue pour que celle-ci soit automatiquement activée lors de chaque visite. C'est aussi un cookie qui est utilisé pour empêcher une personne de participer plusieurs fois à un sondage en ligne. On envoie une information “**a déjà voté**” vers le browser qui ensuite, à chaque nouvelle tentative de participer au scrutin renvoie “**a déjà voté**” vers le serveur, lequel ne l'autorise donc pas à voter une seconde fois.

Ces données doivent rester très petites, car elles doivent être renvoyées lors de chaque requête vers le serveur (c'est à dire lors du téléchargement de chaque nouvelle page ou de chaque nouvelle image). Comme les données personnelles sont de plus en plus grandes on a dû trouver une solution permettant de les stocker directement sur le serveur. Cela est fait en utilisant un cookie unique contenant un nombre, appelé *sessionID*, qui permet d'accéder à toutes les données concernant le client qui sont conservées sur le serveur. Par exemple les sites de commerce électronique utilisent un *sessionID* pour conserver les articles insérés dans le panier virtuel du client. Un *sessionID* est aussi utilisé pour identifier le visiteur d'un site qui est protégé par un mot de passe. Ce sont donc les cookies qui sont utilisés pour récolter les informations les plus privées sur les internautes.

Il est très important de garder les *sessionIDs* petits, car si 10 à 20 % du trafic est généré par ces informations, c'est une grande quantité d'argent qui s'en va. Chaque Mo de donnée transférée coûte environ 0,1\$. Écrire de tels programmes est difficile, mais il existe maintenant des bibliothèques qui font cela automatiquement.