Université
franco-allemande
Deutsch-Französische
Hochschule

Introducing Privacy In Current Web Search Engines
by Albin Petit

Résumé

Au cours des dernières années les progrès technologiques permettant de collecter, stocker et traiter d'importantes quantités de données pour un faible coût, ont soulevés de sérieux problèmes concernant la vie privée. De nombreux services web (ex.: Facebook, Google) ont maintenant la possibilité de créer un profil de leurs clients pour par exemple, leur offrir de la publicité ciblée. La révélation par Edward Snowden en mai 2013 d'un programme de surveillance massif et généralisé opéré par la NSA démontre que les services secrets collectent et exploitent massivement les données personnelles. Ainsi, de nombreuses entités espionnent quotidiennement les interactions des utilisateurs, alors qu'une majorité de ces utilisateurs n'en a toujours pas conscience. Cette ignorance s'explique en partie par le fait que ces services web ne permettent pas à leurs utilisateurs de consulter les données qu'ils collectent. De plus, les conditions générales d'utilisation n'informent pas précisément l'utilisateur sur la nature des données collectées ainsi que sur l'objet de leur collecte.

La protection de la vie privée concerne de nombreux domaines, en particulier les sites internet fréquemment utilisés comme les moteurs de recherche (ex.: Google, Bing, Yahoo!). Ces services permettent aux utilisateurs de retrouver efficacement du contenu parmi une quantité grandissante de données publiées sur internet. La pertinence de leurs recommandations s'explique par l'exploitation des données personnelles des utilisateurs. Par conséquent, les moteurs de recherche sont susceptibles de collecter et stocker des données sensibles d'utilisateurs (ex. : leurs centres d'intérêt, leurs opinions politiques et religieuses, leurs conditions de santé). Dans ce contexte, développer des solutions pour permettre aux utilisateurs d'utiliser ces moteurs de recherche tout en protégeant leurs vies privées est devenu très primordial.

Dans cette thèse, nous introduirons SimAttack, une attaque contre les solutions protégeant la vie privée de l'utilisateur dans ses interactions avec les moteurs de recherche. Cette attaque vise à retrouver les requêtes initialement envoyées par l'utilisateur en exploitant d'anciennes requêtes non protégées précédemment envoyées par l'utilisateur. Nous utiliserons cette attaque pour analyser la robustesse de trois mécanismes de protection représentatifs des solutions existantes. Nous avons montré que ces trois mécanismes ne sont pas satisfaisants pour protéger la vie privée des utilisateurs.

Par conséquent, nous avons développé PEAS, un nouveau mécanisme de protection qui améliore la protection de la vie privée de l'utilisateur (en particulier par rapport à SimAttack). Cette solution repose sur deux types de protection: cacher l'identité de l'utilisateur et masquer sa requête. La première protection chiffre et envoie les requêtes via une succession de deux serveurs tandis que la deuxième protection masque la requête de l'utilisateur en la combinant avec des fausses requêtes. La difficulté majeure de notre approche est de générer des fausses requêtes réalistes. Nous avons résolu ce problème en générant des requêtes qui auraient pu être envoyées par d'autres utilisateurs du système.

Pour finir, nous présenterons des mécanismes permettant d'identifier la sensibilité des requêtes. Notre objectif est d'adapter les mécanismes de protection existants pour protéger uniquement les requêtes sensibles, et ainsi économiser des ressources (ex.: CPU, mémoire vive). En effet, une requête banale sur une recette de gâteau n'a pas besoin de la même protection qu'une requête sur une infection liée au VIH. Nous avons développé deux modules pour identifier les requêtes sensibles et nous avons déployé ces modules sur des mécanismes de protection. Nous avons établi qu'adapter des mécanismes de protection améliore considérablement leurs performances.