

Les empreintes de nos navigateurs nous identifient — et si on brouillait les pistes ?

Depuis un an ou deux une conscience floue de l'ubiquité du pistage est perceptible dans les conversations sous la forme de brèves remarques fatalistes :

— De toutes façons on est espionnés, alors...

— Ils peuvent savoir tout ce qu'on fait sur Internet, hein...

et chez les Dupuis-Morizeau :

— J'ai réservé la location moins cher avec le PC du boulot pour exactement la même chose... Faut croire que depuis qu'on fait des achats en ligne ils se figurent qu'on pourrait payer plus...

Vous qui savez, ne vous moquez pas trop vite : mieux vaut cette perception diffuse qu'une certitude naïve d'être protégé (« ah tu vois le petit s derrière http ça veut dire que c'est sécurisé ce site » hum) ou pire une ignorance dangereuse.

Certes le chemin est long encore avant de pouvoir susciter une plus large prise de conscience puis une évolution des comportements et des usages dans notre vie numérique. C'est à quoi Framasoft s'efforce de mener par étapes avec la campagne longue durée Dégooglisons Internet. Mais bien d'autres initiatives surgissent semaine après semaine, et nous sommes ravis de leur donner un écho. Tel est le cas aujourd'hui avec le projet amiunique (non, ce n'est pas un site de rencontre pour trouver un seul ami pour la vie : Am I unique? signifie : « Suis-je unique ? »).

Nous sommes pistés, traqués, espionnés, piégés, profilés... d'accord, tout le monde en a désormais « entendu parler ». Mais que sait-on de nous au juste lorsqu'on va jeter coup d'œil sur un site web ?

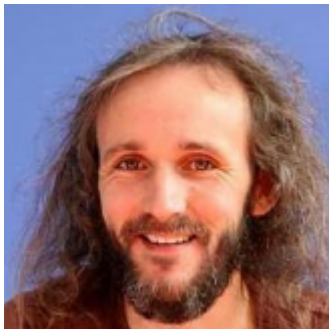
Il existe déjà des moyens d'en avoir une idée, par exemple dans Firefox, de nombreux outils de développement proposent des fonctionnalités pour voir ce qui se passe sous les jupes de la connexion. Mais le site amiunique.org en facilite l'accès et explique de quoi il retourne.

On découvre ainsi que cette quantité importante d'informations permet le plus souvent de faire de vous un utilisateur « unique »

— Ah j'en vois qui se rengorgent de n'être pas comme tous les autres et qui sont tout fiers de se distinguer du supposé troupeau des internautes...^[1]

*Mais ne vous réjouissez pas trop vite, écoutez un peu ce que nous dit **Benoît Baudry** et voyez comment tout au contraire se fondre dans une masse d'empreintes numériques similaires pourrait être un moyen de ne plus être unique, donc de risquer moins d'être pisté !*

— Bonjour, pouvez-vous vous présenter brièvement ?



Je suis chercheur depuis 2004 dans le domaine du génie logiciel. J'ai beaucoup travaillé sur des outils et algorithmes pour améliorer les activités de test logiciel dans le domaine des systèmes embarqués. Cette activité s'est ensuite élargie pour traiter de problèmes d'analyse de programme. Depuis février 2013 je coordonne un projet de recherche européen (DIVERSIFY) sur le thème de la diversification du logiciel

pour faire des applications web plus robustes.

Dans ce contexte, nous avons démarré une activité sur le *browser fingerprinting*, c'est-à-dire les empreintes, les traces numériques que laissent nos plateformes quand nous nous connectons avec nos navigateurs.

La petite équipe qui se consacre à amiunique.org et aux technologies *anti-fingerprint* comprend un doctorant et bénéficie du conseil scientifique et technique de deux chercheurs en sécurité et génie logiciel.

Il n'est pas si fréquent qu'un labo de recherche universitaire entame un projet qui peut toucher un assez vaste public et potentiellement l'ensemble des utilisateurs du Web. Alors pourquoi s'intéresser au fingerprinting ?

Tous nos travaux de recherche sont inspirés par des problèmes réels rencontrés lors de la construction de systèmes logiciels. La nouveauté dans ce cas tient sans doute au fait que le problème ne touche pas seulement les développeurs de logiciels, mais que tout internaute peut se sentir concerné.

Dans le cadre de DIVERSIFY, nous étudions la diversité logicielle et la manière dont elle peut être exploitée, voire amplifiée par des procédures automatiques. L'idée de ce projet c'est que l'augmentation de la diversité logicielle augmente également la résilience et la sécurité des applis webs.

Par ailleurs, nous avons découvert que dans le cas du *browser fingerprinting* la très grande diversité et la richesse de l'écosystème web sont devenues un problème pour la vie privée : cet écosystème très très riche permet à chacun de disposer d'une plateforme qui correspond parfaitement à ses envies et à son confort, mais c'est ce même phénomène qui rend chaque plateforme unique et donc traçable. La question qui nous a alors intéressés est la suivante :

Peut-on exploiter la diversité qui est à l'origine du problème de fingerprint pour lutter contre le traçage, en modifiant automatiquement et régulièrement la plateforme logicielle d'un internaute ?

La première étape pour répondre à cette question consiste à étudier de près la diversité des plateformes qu'on peut trouver sur le web, ce qui nous a amené à développer le site amiunique.org

Bon alors de quoi s'agit-il au juste ? Pourquoi devrais-je me demander si je suis unique ? C'est inquiétant d'être unique sur le Net ?

Une première remarque à propos de l'unicité : quand le site affiche « *Êtes-vous unique ? Oui !* », ce qu'il faut comprendre c'est que la plateforme logicielle (navigateur, OS, polices, plugins, etc.) que vous avez utilisée pour aller sur le site est unique parmi toutes celles observées jusqu'à présent.

Un clic pour essayer...



The screenshot shows a website interface with a green dotted border. At the top, the text reads "Apprenez à quel point votre navigateur est unique" and "Aidez-nous à analyser la diversité des navigateurs web". Below this is a dark blue button with the text "Voir l'empreinte de mon navigateur" and a mouse cursor pointing at it. At the bottom, there is a disclaimer: "Si vous cliquez sur ce bouton, nous ne collecterons que des données anonymes et nous installerons un cookie sur votre navigateur pour 4 mois. Plus de détails sont disponibles dans la [politique de stockage et de protection des données](#)."

et voici le résultat :

Êtes-vous unique ?

Oui ! (Vous pouvez être tracé !)

45.92 % des navigateurs observés sont Firefox, comme le vôtre.

0.29 % des navigateurs observés sont Firefox 37.0, comme le vôtre.

18.40 % des navigateurs observés sont installés sur Linux, comme le vôtre.

17.66 % des navigateurs observés sont réglés en "fr", comme le vôtre.

34.22 % des navigateurs observés sont réglés sur le fuseau horaire UTC+1, comme le vôtre.

Cependant, la combinaison de tous les éléments constituant votre empreinte est unique parmi les 47977 empreintes déjà récoltées. Cliquez ici pour savoir pourquoi.

Voir plus de détails

Voir graphiques

Quand vous visitez un site web, celui-ci peut facilement récolter un grand nombre d'informations à propos de votre plateforme. Certaines de ces informations sont systématiquement fournies au site web par le navigateur (*user agent* et en-têtes http), et pour le reste le serveur peut interroger les interfaces Flash, JavaScript et toute autre interface offerte par le navigateur et ses plugins.

Une longue liste de données que l'on peut récolter... (cliquer pour agrandir)



Le problème d'avoir une plateforme unique c'est que les sites marchands peuvent personnaliser leur contenu en fonction de la plateforme (pratiques tarifaires en fonction des clients ou publicités ciblées). Par exemple, les sites peuvent varier les prix en fonction de la localisation de l'internaute, déduite approximativement par l'adresse IP (voyez cet exemple), ou en fonction du système d'exploitation (un autre exemple). Une étude a montré que le site de voyages *Orbitz* propose des prix plus élevés aux utilisateurs de Mac OS (exemple d'Orbitz).

Mais ce que vous proposez de faire, ça commence bien comme du pistage non ? Vous voulez qu'on vous donne toutes les informations sur notre navigateur pour votre recherche, vous êtes sympathique avec votre bonne tête de nerd barbu mais comment puis-je avoir confiance ? Et puis vous allez faire quoi de cette bientôt énorme base de données récoltées,

mmh ?

La politique de confidentialité d'amiunique.org établit très clairement ce qui advient des données récoltées : elles sont stockées sur une machine sécurisée et nous les utilisons à des fins d'analyse statistique. Par ailleurs, nous ne récoltons absolument aucune information qui nous permette de relier une empreinte à une personne, les données sont donc anonymes par construction. Les analyses statistiques d'empreintes réelles sont essentielles pour établir des procédures qui modifient automatiquement la plateforme d'un internaute de manière réaliste. Dans cet objectif, 2 défis à relever : réussir à modifier la plateforme de manière transparente, mais suffisamment pour changer l'empreinte ; modifier la plateforme tout en la faisant ressembler à une vraie plateforme pour éviter d'être repéré comme un usager qui essaie de se cacher.

Quelle est selon vous la masse critique d'empreintes collectées qui vous permettra de passer à la phase suivante du projet ?

Il est important pour nous de rassembler un maximum d'empreintes pour que le verdict que nous rendons aux visiteurs (oui / non vous êtes unique) soit vraiment significatif. À l'échelle du Web, être unique parmi quelques dizaines de milliers d'internautes, ce n'est pas forcément très impressionnant, alors qu'être unique parmi quelques millions c'est déjà plus surprenant. Pour nos travaux futurs sur le brouillage c'est surtout important de récupérer une très grande diversité d'empreintes (diversité géographique, diversité de navigateur, polices, etc.). Un échantillon étendu et diversifié nous permettra d'identifier quelles sont les caractéristiques les plus discriminantes, qui devront être la cible principale pour le brouillage.



et euh... je ne risque rien si je "triche" en envoyant finalement des données trompeuses ? C'est bien légal ça ?

Il n'y a rien d'illégal à modifier/changer/envoyer des fausses informations. Cependant, deux problèmes peuvent se poser avec les extensions de navigateurs qui renvoient de fausses informations dans le user agent (par exemple User agent switcher).

Le premier vient du fait que les informations que vous envoyez à un serveur sont censées être utilisées pour vous renvoyer une page web qui est parfaitement

adaptée à votre navigateur et à sa configuration. Si vous mentez, la page web peut présenter de nombreux défauts et certaines fonctionnalités peuvent ne pas être opérationnelles.

Le second problème vient du fait que si vous mentez, il se peut que votre empreinte soit incohérente et cela vous rend tout de suite identifiable. Par exemple, sur une des empreintes récoltées sur amiunique, le navigateur annonce dans son *User Agent* qu'il tourne sous Windows alors que le moteur de rendu JavaScript nous indique que la machine en question est un Mac. Cette configuration qui n'est pas censée exister dans la réalité vous rend identifiable car vous serez une des seules personnes au monde à avoir cette configuration incohérente.

Ce n'est donc pas illégal de mentir, mais ça peut engendrer des problèmes à la fois liés à la navigation et au traçage par empreinte.

Par ailleurs, il n'y a rien d'illégal à modifier charger / supprimer des nouvelles polices à chaque lancement de votre navigateur ou à utiliser régulièrement une version différente d'un navigateur.

Nous pensons qu'un avantage de *modifier la plateforme* (plutôt que de renvoyer de fausses informations) c'est qu'on ne prendra pas le risque d'être identifié comme un menteur ou comme une plateforme qui tente d'envoyer des données trompeuses.

C'est astucieux tout ça mais vous ne craignez pas que ceux qui nous pistent trouvent une parade à votre dispositif ?

Certainement et c'est toute la difficulté de tout dispositif pour la vie privée : dès qu'un nouveau dispositif apparaît, il fixe également les nouvelles barrières à franchir pour ceux qui veulent nous pister. Ce que nous espérons surtout c'est que ce genre de mécanisme rende le pistage plus difficile et plus coûteux.

Est-ce qu'on peut espérer que cette expérience "de laboratoire" aura une utilité pour nous les internautes concernés par le pistage de nos empreintes ? Pour votre équipe c'est un intéressant objet de recherche, mais sous quelle forme finale pourrait aboutir votre projet pour la famille Dupuis-Morizeau^[2] ?

Nous l'espérons vraiment. Pour l'instant, il faudrait idéalement que la famille Dupuis-Morizeau ait des PC sous Linux, et dans ce cas nous visons une solution

qui puisse être téléchargée et utilisée à la place du navigateur standard. En gros, ça devrait être un conteneur qui isole le navigateur et tous les éléments de plateforme, qui est lancé à chaque fois que quelqu'un veut naviguer sur Internet. Au lancement de ce conteneur, un navigateur, des polices, des plugins, voire quelques éléments de configurations système sont sélectionnés aléatoirement, assemblés et démarrés. L'internaute peut alors naviguer avec une plateforme différente de celle utilisée la dernière fois.

Bon, on y voit un peu plus clair et ce projet a de quoi intéresser. Que peuvent faire les lecteurs de cet article s'ils souhaitent contribuer ?

Distribuer l'URL pour augmenter le nombre d'empreintes c'est déjà une énorme contribution. Toute idée de forums internationaux pour augmenter la diversité des empreintes est aussi tout à fait bienvenue. Enfin, le code du site est *open source* et nous serions ravis d'avoir des contributions à ce projet sous forme de traduction de pages, revue de code (architecture du site, sécurité, etc.) ou toute idée pour mieux valoriser les données auprès des visiteurs.

Maxime Dupuis-Morizet, bêta-testeur pour le projet amiunique...



Liens

- Le site amiunique.org
- Contribuer sur le github
- la foire aux questions et la politique de confidentialité
- l'INSA-Rennes et l'INRIA

Notes

[1] Si vous pensez que les internautes moyens sont Mme Michu, une dinde ménopausée, et son mari tonton Roger, que ses gros doigts de pêcheur de brochet empêchent d'appuyer sur la bonne touche, cet article n'est pas pour vous.

[2] L'authentique Mme Michu a demandé qu'on cesse de la mentionner. Le Framablog a décidé d'employer désormais l'exemple des Dupuis-Morizeau, une sympathique famille recomposée qui vit à Rouen en Normandie.