Le Réseau TOR comme moyen de protection de la vie privée sur internet

Selim Ben Ismail

03 Jan 2020

Résumé

Cet article scientifique, rédigé dans le cadre du cours d'architecture des systèmes d'information de l'Université Libre de Bruxelles, aura pour objet d'appréhender la nature du réseau TOR, d'en comprendre les principes mis en œuvre et d'en évaluer les possibilités actuelles et futures concernant la protection de la vie privée sur internet. L'article est divisé en quatre parties : dans un premier temps, nous introduirons les concepts et enjeux liés à la thématique de l'anonymat sur internet. Dans un second temps, nous explorerons les différents systèmes permettant de protéger les données et ferons un bref état de leur fiabilité ainsi que de leurs limites. Ensuite nous nous attarderons sur le cas controversé de TOR pour en décrire sa nature, ses mécaniques internes et son évolution technique comme idéologique. Finalement, la dernière partie de l'article sera consacrée à une réflexion sur les enjeux que représente la sécurisation des données à caractères personnels dans notre société actuelle.

Introduction

Dans cette ère du web 2.0, où existe une interconnexion quasi permanente des personnes entres elles par le biais des services numériques, l'humain se voit être progressivement représenté par des ensembles de données personnelles. Ainsi on observe un phénomène de documentalisation de la personne qui croît de façon parallèle au développement des technologies de l'information et de la communication. Ces systèmes informationnels alimentés par nos interactions sur les réseaux socio-numériques et nos saisies endogènes dans la toile sont devenus des éléments clés de notre société hyperconnectée. Ressources commerciales, outils de spéculation, instruments de surveillance, substrats de recherche, e.t.c. ; les champs d'application sont sans limites pour qui sait les saisir.

Cependant, ces informations d'apparence anodines n'en restent pas moins des éléments de notre vie privés, et peuvent être sensibles, préjudiciables, ou simplement fausses. Il convient, dès lors, de se questionner sur la gouvernance de celles-ci, la confidentialité qui les entoure et l'impact que leur exploitation pourrait avoir sur nos libertés individuelles.

Heureusement, cette collecte industrialisée de nos données à caractères personnels n'est pas une fatalité : d'une part, celle-ci est encadrée par un certain nombre de lois et de clauses1, d'autre part, il existe un certain nombre de dispositifs de sécurité permettant de chiffrer nos informations ainsi que le contenu de nos échanges au travers des réseaux ou de bloquer des requêtes jugées

1 Définition et enjeux relatifs à la protection de l'identité

Les données à caractères personnels

Avant d'entrer plus dans le détail, il est nécessaire de définir concrètement cette notion

de “donnée à caractère personnel”, couramment abrégée en “données personnelles” ou “DCP”. La définition légale, que l’on peut retrouver sur le site officiel de la Commission Européenne, est entérinée comme telle :

> "Les données à caractère personnel sont des informations se rapportant à une personne vivante identifiée ou identifiable. Différentes informations, dont le regroupement permet d’identifier une personne en particulier, constituent également des données à caractère personnel. Des données à caractère personnel qui ont été rendues anonymes, chiffrées ou pseudonymisées, mais qui peuvent être utilisées pour identifier à nouveau une personne constituent toujours des données à caractère personnel et sont couvertes par le RGPD. Les données à caractère personnel rendues anonymes de telle manière que la personne ne soit pas ou plus identifiable ne constituent plus des données à caractère personnel. Pour qu’une donnée soit véritablement rendue anonyme, le processus d’anonymisation doit être irréversible. [...]"

Rentre dans cette définition, les éléments tels que le nom, le prénom, les adresses de résidence ou de courriel, le numéro de carte d’identité, les données de géolocalisation, les adresses IP et MAC, les cookies, ou toutes autres données détenues par un hôpital ou un médecin, qui permettraient d’identifier de manière unique une personne.

> "Il faut insister sur le fait que l’«identification» ne désigne pas simplement la possibilité de retrouver le nom et/ou l’adresse d’une personne, mais inclut aussi la possibilité de l’identifier par un procédé d’individualisation, de corrélation ou d’inférence"

C’est en effet dans la conjointure des big data aux algorithmes de datamining que des données - à première vue désuètes - révèlent une dimension pernicieuse. L’affaire AOL, est un exemple typique de cette surestimation du pseudonymat dans la protection de la vie privée. En août 2006 une équipe de chercheurs d’AOL diffusa publiquement une base de données contenant les saisies de recherche effectuées par près de 658 000 internautes au cours des mois de mars, avril et mai 2006, où l’identifiant des utilisateurs avait remplacé par une suite de caractères numériques. Très rapidement, la base de données fut rediffusée par des tiers sur des serveurs miroirs rendant son retrait de la toile presque impossible. Et malgré l’immense océan que représentait les données à traiter - il était question de quelques 20 millions de requêtes enregistrées - de nombreux internautes purent être identifiés par le recoupement d’informations au travers des différentes sources de données. On y trouve par exemple, l’utilisateur n°4417749, identifié en la personne de Thelma Arnold, âgée alors de 62 ans, vivant à Lilburn dans l’état de Georgie des États-Unis qui aurait interrogé le moteur de recherche AOL sur : les célibataires de 60 ans, les tremblements de la main, les effets de la nicotine et les chiens qui urinent partout. Si ce cas peut paraître anecdotique, le problème que peut poser une mauvaise gestion des données utilisateurs quant à la protection de leur vie privée, y apparaît très clairement et il n’est pas surprenant que cet évènement ait susciter la colère des citoyens américains.

Les cookies

Un cookie est un petit fichier texte généré par les serveurs HTTP et envoyé aux clients lorsque ceux-ci s’y connectent pour la première fois. Ce fichier est stocké en local dans le cache du navigateur internet et sera renvoyé chaque fois que le client voudra interroger ce même serveur HTTP.

Il existe différents types de cookies et nous n’entrerons pas plus dans le détail

de leur fonctionnement, mais leur rôle initial est de permettre à un serveur de “reconnaître” un client et de fluidifier la navigation de ce dernier en conservant des informations spécifiques à l'utilisateur. Ce sont les cookies qui permettent des fonctionnalités comme l'identification automatique, le maintien d'une session, l'exécution d'un contenu flash ou la mémorisation de nos préférences sur un site pour une reconnexion ultérieure. Malheureusement, ce sont eux aussi qui permettent le profilage de l'utilisateur et le tracking publicitaire. Les informations principales contenues dans un cookie sont : un identifiant propre à l'internaute, le domaine de provenance du cookie, sa date de création et un contenu lié à l'utilisateur (identifiant de session, login, mot de passe, e.t.c) . En inspectant le fichier, il est donc possible d'apprendre que l'utilisateur X a visité une page web Y à l'instant T, plaçant les cookies dans le champ de définition des données à caractères personnels et sous couvert du RGDP.

2 Les dispositifs de sécurité des données

2.1 Le mode de navigation privée

L'outil de navigation privée intégré aux navigateurs web n'est malheureusement qu'un leurre. Sa dénomination, presque malhonnête, tend à rendre cet outil populaire et à le surcoté dans l'imaginaire collectif. Pourtant, ce mode de navigation ne protège d'aucunes manière l'internaute d'une quelconque forme d'identification ou de pistage qui aurait eu lieu lors d'une navigation “normal”. Le mode de navigation privée empêche simplement l'enregistrement local de l'historique de navigation et supprimes les cookies HTTP stockés dans le cache du client à la fermeture de la session. Mais cela n'affectera en rien les données perçues et enregistrées du coté serveur.

2.2 Les serveurs proxy

Les serveurs proxy, qu'on appelle aussi serveurs mandataires, sont des dispositifs que l'on retrouve fréquemment comme éléments de sécurité aux sorties des réseaux locaux. Le principe de fonctionnement est relativement simple : le serveur proxy fait office d'intermédiaire entre deux hôtes – souvent l'un situé à l'intérieur d'un réseau local tandis que l'autre est sur l'internet mondial – relayant les requêtes qu'il reçoit et analysant le trafic qui le traverse. De cette façon, il va mettre à disposition du client ses propres ressources (mémoire cache, compression des données, etc) afin d'accélérer sa navigation et filtrer les contenus non-désirables comme les pubs ou les contenus malveillants.

Paradoxalement, si le serveur proxy peut être utilisé comme un outil de censure et de contrôle en restreignant l'accès des clients du réseau sous-jacent à des sites spécifique ou en bloquant certains contenus, il pourra aussi être utiliser comme un outil d'émancipation. En effet, lorsque nous accédons au grand World Wide Web par l'entremise d'un serveur proxy, nous ne naviguons plus avec notre adresse IP, mais avec celle du serveur proxy. Et c'est là, tout l'intérêt de ce dispositif dans la protection des données à caractère personnels. Car dès lors que notre adresse IP, notre identifiant numérique principal, est anonymisé, il est significativement plus laborieux de pister nos activités sur la toile. Qui plus est, nous masquons au passage les informations qui y étaient liées, comme notre localisation géographique par exemple. Pour illustrer ça : supposons une personne résidente d'un certain pays souhaitant se connecter à un service spécifique ; que, pour une quelconque raison politico-juridique, la liaison à celui-ci est bloquée dans sa zone. En passant par un serveur proxy situé dans un autre pays, cette personne pourra alors contourner la restriction géographique et joindre le service qu'elle convoitait.

Toute foi, si le proxy intègre des possibilités intéressantes pour protections de nos

données à caractères personnels, il n'est pas sans failles : il souffre de deux défauts non-négligeables.

Le premier est que la liaison entre le client et le serveur proxy n'est pas protégée et peut être facilement le sujet d'une attaque. Le second est que si l'adresse IP se voit être masquée, ce n'est pas le cas du contenu des requêtes envoyées sur le réseau. Et finalement, la plus grande faille du proxy concernant la protection de la vie privée, c'est le proxy lui-même. Il est capital de garder à l'esprit que le serveur dispose toutes les informations que nous voulions initialement cacher et qu'il peut potentiellement les enregistrer ainsi que toutes les autres données confidentielles passant par lui (identifiants, coordonnées banquaires, etc). Il convient donc de s'assurer de la bienveillance des titulaires du serveur proxy mais aussi de la qualité de ses défenses.

2.3 Les VPN

2.4 Les MixNet

3 Le réseau TOR

3.1 Nature du réseau

3.2 Principe de fonctionnement

3.3 Évolution

3.4 Positionnement Ethique

4 Conclusion

Références

1. « Article 8 - Protection des données à caractère personnel ». Agence des droits fondamentaux de l'Union européenne, 25 avril 2015, https ://fra.europa.eu/fr/charterpedia/article/8-protection-des-donnees-caractere-personnel.
2. « À quoi correspondent les données à caractère personnel? » Commission européenne - European Commission, https ://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_fr. Consulté le 30 décembre 2019.
3. Barthelemy, Brieuc. L'anonymat dans le reseau Tor (The Second-Generation Onion). p. 25.
4. Bénédicte, REY. L'insécurité numérique au quotidien : de la régulation quotidienne aux logiques d'alertes. p. 12. Bosqué, Camille. « Tor, la face chiffrée d'Internet ». Vacarme, vol. N° 69, n 4, octobre 2014, p. 79-98. www.cairn.info, https ://www.cairn.info/revue-vacarme-2014-4-page-79.htm.
5. « Cookie (informatique) ». Wikipédia, 9 décembre 2019. Wikipedia, https ://fr.wikipedia.org/w/index.php ?title=Cookie_(informatique)oldid=165278416.
6. DE REYNAL, Denis. Présentation sur les VPN.
7. Dingledine, Roger, et al. Tor : The Second-Generation Onion Router : Defense Technical Information Center, 1 janvier 2004. Crossref, doi :10.21236/ADA465464.
8. Edward Snowden - interview partie 1 [sous-titres français]. YouTube, https ://www.youtube.com/watch ?v=8iVojVSqs74. Consulté le 27 décembre 2019.
9. Edward Snowden - Interview partie 2 [sous-titres français] - « Ils diront que j'aide nos ennemis ». YouTube, https ://www.youtube.com/watch ?v=oC8OOkUT1NU. Consulté le 27 décembre 2019.
10. FILIOL, Eric. Statistical and combinatorial analysis of the TOR routing protocol : structural weaknesses identified in the TOR network. Consulté le 22 novembre 2019.
11. Fonctionnement de la navigation privée dans Chrome - Ordinateur - Aide Google Chrome. https ://support.google.com/chrome/answer/7440301. Consulté le 2 janvier 2020.
12. Gambs, Sébastien. Réseaux de Communication Anonyme. p. 39.
13. Huber, Markus, et al. « Tor HTTP Usage and Information Leakage ».
14. Communications and Multimedia Security, édité par Bart De Decker et Ingrid Schaumüller-Bichl, Springer, 2010, p. 245-55. Springer Link, doi :10.1007/978-3-642-13241-4_22.
15. Informations vendues sur le darknet : «Une problématique constante des services de renseignements». YouTube, https ://www.youtube.com/watch ?v=7HkWhqbOSks. Consulté le 29 décembre 2019.
16. LE GROUPE DE TRAVAIL SUR LA PROTECTION DES PERSONNES À L'ÉGARD DU TRAITEMENT DES DONNÉES À CARACTÈRE PERSONNEL.
17. Lecomte, Romain. « L'anonymat comme «art de résistance». Le cas du cyberespace tunisien ». Terminal. Technologie de l'information, culture société, n 105, octobre 2010. journals.openedition.org, doi :10.4000/terminal.1862.

18. Pardonnez-moi - L'interview d'Edward Snowden. YouTube, https ://www.youtube.com/watch?v=e94nv7zca-k. Consulté le 27 décembre 2019.
19. Paveau, Marie-Anne. « Éthique du discours numérique ». Línguas e Instrumentos Linguísticos (Brésil), vol. 37, 2016, p. 177-210. HAL Archives Ouvertes, https ://hal.archives-ouvertes.fr/hal-01423473.
20. Perrat, Jean-François. « Un «Deep / dark web»? Les métaphores de la profondeur et de l'ombre sur le réseau Tor ». Netcom. Réseaux, communication et territoires, n 32-1/2, décembre 2018, p. 61-86. journals.openedition.org, doi :10.4000/netcom.3134.
21. Pierre, Julien. Internet et les données à caractère personnel : traitement, enjeux et gouvernance. p. 14.
22. Pillot, Guillaume. Anonymat et vie privée sur internet. 2018. corpus.ulaval.ca, https ://corpus.ulaval.ca/jspui/handle/20.500.11794/32469.
23. Piolle, Guillaume. Anonymat, anonymisation, désanonymisation. —. Outils informatiques pour la protection de la vie priv´ee. p. 6.
24. Rumors of Tor's compromise are greatly exaggerated | Tor Blog. https ://blog.torproject.org/rumors-tors-compromise-are-greatly-exaggerated. Consulté le 29 décembre 2019.
25. Sangeetha, K., et K. Ravikumar. « A Novel Traffic Dividing and Scheduling Mechanism for Enhancing Security and Performance in the Tor Network ». Indian Journal of Science and Technology, vol. 8, n 7, p. 689–689. cibleplus.ulb.ac.be, doi :10.17485/ijst/2015/v8i7/62882. Soltani, Ashkan, et al. « Flash Cookies and Privacy ». SSRN Electronic Journal, 2009. DOI.org (Crossref), doi :10.2139/ssrn.1446862.
26. Tréguer, Félix. Anonymat et chirement, composantes essentielles de la liberté de communication. p. 28.
27. Yannakogeorgos, Panayotis. « Internet Governance and National Security ». Strategic Studies Quarterly, vol. 6, n 3, p. 102–125.