La vie privée sur l'Internet à l'ère du RGPD

Stéphane Bortzmeyer stephane+jdll@bortzmeyer.org

JDLL, 24 mars 2018

Plan

Chez les utilisateurs

- Ignorance (« Ils ne peuvent quand même pas retrouver mes messages parmi 500 millions d'utilisateurs » « On m'a juré que les données étaient cryptées et anonymisées »),
- Confiance (« Je n'ai rien à cacher, et je pense que Macron veut notre bien »)
- Fatalisme (« De toute façon, on est tous fichés partout tout le temps »)

Chez les administrateurs système

- On stocke tout, dans le doute,
- On fait des sauvegardes,
- On ne détruit jamais rien (ou alors par accident),
- On sécurise. . . moyennement,
- **Il n'y a pas que les GAFA qui abusent des données !**

Chez les décideurs ou juristes

- Pas toujours au courant de ce qui est stocké
- Les listes de fichiers de données sont rarement complètes (journaux, pcap. . .)

Chez les développeurs

- Des formulaires et des bases de données avec trop d'informations

Plus récemment

- Avant, on gardait trop de données par négligence
- Maintenant, « les données sont le pétrole du XXIe siècle »
- Tout le monde croit que *Big Data* et *Business Intelligence* vont rapporter plein de brouzoufs
- Une bonne (?) raison de garder plein de données

Trois cas

- Le GAFA, qui capte plein de données personnelles, mais a de grosses équipes d'avocats et est probablement compatible avec le RGPD, et se sert effectivement des données,
- la startup FrenchTech qui récolte plein de données, ne sait pas pourquoi et ne les utilise pas (*Dark Data*),
- le libriste moyen qui récolte des données parce que c'est le comportement par défaut sur Debian.

Et le risque de piratage

- « Notre base d'utilisateurs est sécurisée. Nous utilisons du cryptage de haut niveau. »
- La vérité est que votre base sera piratée, tôt ou tard,
- Et vous ne le saurez même pas,
- Bien sûr, il faut prendre toutes les précautions possibles pour diminuer ce risque (ne pas laisser une copie de la base sur un MongoDB grand ouvert),
- Mais il faut aussi raisonner sur ce qui arrivera en cas de malheur,
- Colmater quelques failles de sécurité au fur et à mesure n'est pas une approche suffisante.

Mais il n'y a pas que le risque de piratage

- Si vous récoltez des données personnelles, « comment les protéger ? » ne devrait être que la troisième question,
- « Est-ce que je devrais les récolter ? » sera la deuxième question,
- La **première question** est plutôt « qu'est-ce que les pires salauds feraient de ces données s'ils pouvaient mettre la main dessus ? »

Rappels

- Réglement européen (donc s'applique partout « pareil »),
- S'applique à partir du 25 mai 2018,
- « Le traitement des données à caractère personnel devrait être conçu pour servir l'humanité. »

Rien de nouveau sous le soleil ?

- En France, des gens s'affolent sur des obligations du RGPD. . . qui existent depuis la loi Informatique & Libertés de 1978 !
- Le principe de minimisation existe depuis longtemps,
- L'obligation de consentement, de justification de la récolte, existait déjà,
- Les droits d'accès, de suppression et de rectification aussi.

Vraies nouveautés RGPD, pour la France

- Attention, l'auteur n'est pas juriste (mais il voudrait être Tom Hanks dans « Le pont des espions »),
- Non-territorialité (dès qu'on gère des données de résidents européens, on doit respecter le RGPD),
- Démontrer le respect de la protection des données personnelles (plus de « ne vous inquiétez pas, on est sécurisé *military-grade* »),
- Droit à la portabilité des données,
- Amendes plus dissuasives,
- Obligation de notification des failles de sécurité,
- Renforcement des obligations d'information et de transparence.

En fait, le plus gros intérêt du RGPD est de servir de **piqûre de rappel**.

Tout le monde aime le RGPD

- Les spams « Soyez RGPD compliants en quinze jours » ne cessent pas,
- Risque de « *privacy-washing* ».

Enfin, non, pas tout le monde

"L'Europe risque de devenir le tiers monde du numérique à cause de sa législation" pour le PDG de Sigfox

Publié le 20 décembre 2017 par La Revue du Digital dans Evènement avec Aucun commentaire

Bon, il y a toujours des râleurs

Emmanuel Torregano
@ZaraA

Suivre

Bon j'ai lu, la #RGPD est donc calibrée pour s'assurer que la france ne prendra jamais le train de l'innovation et de la disruption. Un règlement fondamentalement réactionnaire et protectionniste. Tellement symptomatique.

11:21 - 21 févr. 2018

4 Retweets 7 J'aime

11 4 7

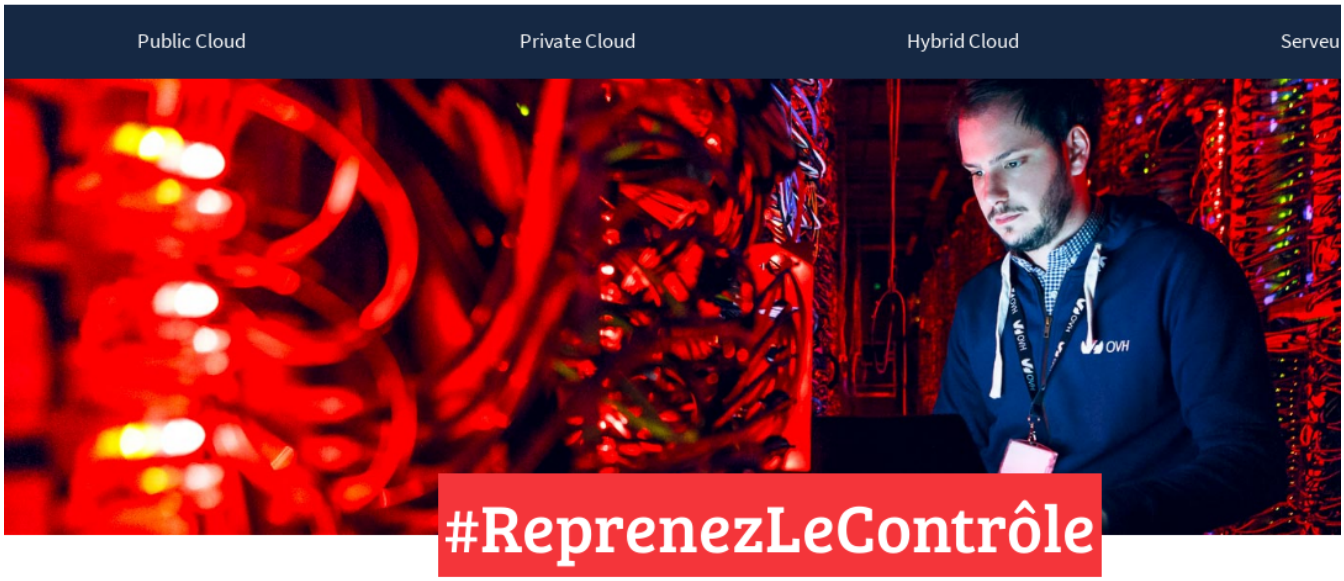
Tweeter votre réponse

Emmanuel Torregano @ZaraA · 22 h

Quand même, le RGPD, c'est bon pour le commerce

OVH Innovation is Freedom

Partagez sur les réseaux #ReprenezLeContrôle



La domination de quelques entreprises sur Internet est source de préoccupations. À juste titre. L'omniprésence des services numériques dans notre quotidien — au travail comme dans notre vie personnelle — fait de la question des

s'intéresser à l'endroit où ces données sont stockées et traitées : le Cloud*.

Les données ne sont pas le pétrole, elles sont un risque pour vous !

- Avoir beaucoup de données personnelles sur ses disques durs n'est pas un avantage,
- C'est une responsabilité, et un risque.

Les protocoles sont trop bavards

- Exemple, DHCP, qui crie à tout le monde le nom de la machine « je suis l'iPhone de Jean-Kevin ! » et son ancienne adresse « et j'étais précédemment à l'université de Grenoble ! »,
- Le DNS transmet à plein de gens votre intérêt vis-à-vis de `alcooliques-anonymes.org`,
- HTTP envoie bien trop d'informations (il n'y a pas que les *cookies*).

Surtout, il faut récolter moins de données

On accepte souvent :

Prénom :

Nom :

Vous êtes : Homme

Homme
Femme

Tiens, pourquoi pas :

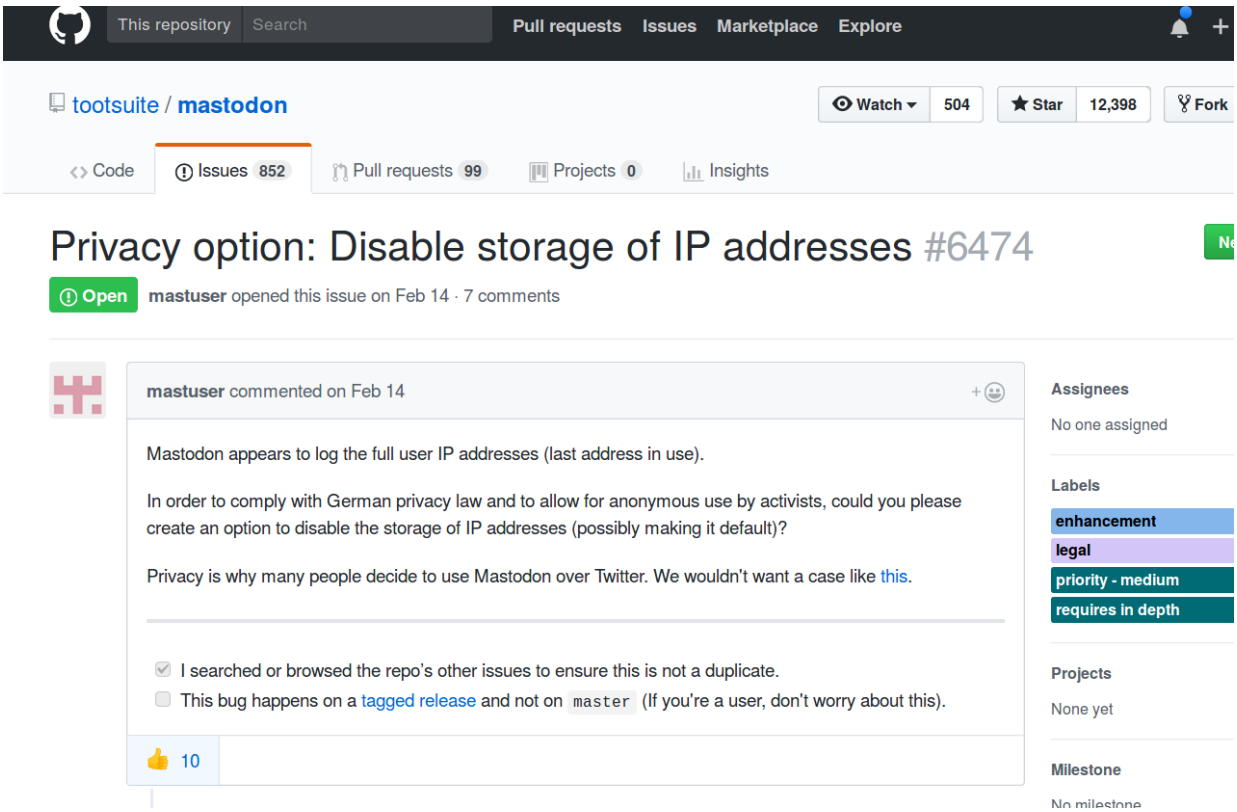
Prénom :

Nom :

Vous êtes : Noir

Noir
Jaune
Blanc

Perdre l'habitude de tout garder par défaut



Le cas des journaux

- Fichiers conservant l'historique des activités de la machine,
- Souvent bourrés d'informations personnelles,
- Peu de professionnels y pensent (ce n'est pas vu comme une base de données),
- Recommandation 1 : vérifier qu'ils ne sont pas gardés trop longtemps,
- Recommandation 2 : séparer les données **temporaires**, très complètes, et les données permanentes, agrégées et brouillées.

« Anonymiser » les données

- **Attention**, ce n'est pas parce qu'il n'y a pas de nom dans les données qu'elles sont anonymes,
- Le sens grand public du terme « anonymisation » est très trompeur, il s'agit juste en général de **pseudonymisation** (il y a toujours un identificateur unique, parfois implicite),
- Et les techniques modernes permettent très souvent de désanonymiser.

Brouiller les données

- La technique la plus efficace est en général de diminuer la quantité d'informations (développeurs, ne gardez pas tous les bits),
- Adresses IP : ne gardez que des préfixes /16 (IPv4) et /32 (IPv6),
- Exemple, si vous stockez une localisation, limitez la précision,
- Attention, un nombre de kilomètres constant n'est pas une bonne idée : à Lyon, un kilomètre suffit à brouiller, mais pas à la campagne,
- Autre exemple, sur une date de naissance, ne gardez que l'année (c'est suffisant pour des statistiques).

Problèmes opérationnels

- Évidemment, moins de données peut être un problème,
- Si on garde les journaux une semaine, on aura du mal à enquêter sur un incident qui s'est produit 10 jours auparavant,
- Si on met en œuvre les recommandations du RFC 7844 sur DHCP, le travail de l'administrateur système sera plus difficile,
- La sécurité est toujours un compromis,
- Pas de miracle, les seules techniques efficaces de protection de la vie privée diminuent les données.

Problèmes légaux

- Les lois peuvent être contradictoires,
- Concilier le RGPD avec les lois « état d'urgence »
- Cas d'une réquisition judiciaire (Quadrature du Net avec leur instance Mastodon en octobre 2017) : respecter l'arrêt Tele2 Sverige AB (C-203/15) rendu le 21 décembre 2016 par la Cour de justice de l'Union européenne (14 jours maximum) ou bien l'article 6, II, de la loi du 21 juin 2004 pour la confiance dans l'économie numérique (un an minimum) ?