Des métadonnées Twitter…

S’il est de notoriété publique que nos données personnelles sont enregistrées et utilisées par les G.A.F.A.M., il est en revanche moins connu que certaines de ces données sont utilisables par tout le monde. Et c’est bien là le point faible de toute campagne de prévention : on a beau dire que nos données sont utilisées, il est peu fréquent que nos paroles soient illustrées.

x0rz publie sur son blog un billet qui illustre parfaitement ce problème. En effet, il a écrit un petit script Python (moins de 400 lignes de code) qui récupère et synthétise les métadonnées Twitter, accessible par n’importe qui.

Ce billet ouvre deux perspectives :

- Concernant le harcèlement numérique : certes ces données sont publiques, mais il faut tout de même quelques capacités en programmation pour les exploiter, ce qui n’est pas à la portée de tout le monde. Imaginons qu’apparaissent de plus en plus de programmes grand public permettant d’accumuler et synthétiser ces données. Il deviendra alors plus facile pour un particulier d’identifier et de traquer une autre personne.
- Concernant les métadonnées en général : dans cet exemple, les données analysées restent très basiques (heure et localisation). Nous arrivons toutefois, par l’accumulation et le recoupement, à déduire des informations intéressantes de ces « méta-métadonnées », et à identifier nettement une personne. Imaginons que les métadonnées enregistrées soient plus précises et plus nombreuses, les informations obtenues seraient alors d’une importance et d’une précision inimaginables. Est-ce alors nécessaire de mentionner qu’à la fois les entreprises (ici Twitter) et les agences

gouvernementales ont accès à ce genre de métadonnées ?

Article original écrit par x0rz, consultant en sécurité informatique, sur son blog.

Traduction Framalang : mo, mathis, goofy, valvin, Diane, Moriarty, Bromind et des anonymes

Vous serez surpris par tout ce que vos tweets peuvent révéler de vous et de vos habitudes

Une analyse de l'activité des comptes Twitter

J'utilise Twitter tous les jours. Pour moi qui suis consultant en cybersécurité, c'est de loin un des meilleurs outils pour rester informé des dernières actualités et pour partager des informations qu'on estime pertinentes pour d'autres. Avec la récente investiture de Donald Trump, les dingos de Twitter de la nouvelle administration et l'émergence de groupes de résistance sur Twitter, j'ai décidé de démontrer à quel point il est facile d'exposer des informations révélatrices à partir du compte de quelqu'un d'autre, sans même le pirater.

Métadonnées

Comme tous les réseaux sociaux, Twitter sait beaucoup de choses sur vous, grâce aux *métadonnées*. En effet, pour un message de 140 caractères, vous aurez un paquet de métadonnées, plus de 20 fois la taille du contenu initial que vous avez saisi ! Et vous savez quoi ? Presque toutes les métadonnées sont accessibles par l'API ouverte de Twitter.
Voici quelques exemples qui peuvent être exploités par n'importe qui (pas seulement les gouvernements) pour pister quelqu'un et en déduire son empreinte numérique :

- Fuseau horaire et langue choisie pour l’interface de twitter
- Langues détectées dans les tweets
- Sources utilisées (application pour mobile, navigateur web…)
- Géolocalisation
- Hashtags les plus utilisés, utilisateurs les plus retweetés, etc.
- Activité quotidienne/hebdomadaire

Un exemple d’analyse de tweet (2010, l’API a beaucoup changé depuis).

Tout le monde connaît les dangers des fuites de géolocalisation et à quel point elles nuisent à la confidentialité. Mais peu de gens se rendent compte que tweeter de façon régulière suffit à en dire beaucoup sur vos habitudes.
Prendre séparément un tweet unique peut révéler des métadonnées intéressantes. Prenez-en quelques milliers et vous allez commencer à voir se dessiner des lignes directrices. C’est là que ça devient amusant.

Méta-métadonnées

Une fois qu’on a collecté suffisamment de tweets d’un compte on peut par exemple identifier ceux qui relèvent d’une entreprise (émettant uniquement pendant les horaires de bureau) et même essayer de deviner combien d’utilisateurs interagissent avec ce compte.
Pour prouver ce que j’avance, j’ai développé un script en python qui récupère tous les derniers tweets de quelqu’un, extrait les métadonnées, et mesure l’activité en fonction de l’heure et du jour de la semaine.

Analyse du compte de @Snowden

Snowden a posté 1682 tweets depuis septembre 2015. Comme on peut le voir ci-dessous, il est facile de déterminer son rythme de sommeil (fuseau horaire de Moscou).

Activité du compte Twitter de Snowden

Analyse du compte de @realdonaldtrump

Est-ce que le compte de Donald Trump est géré par plusieurs personnes ? Cette fois en observant le nombre de sources détectées, je vous laisse deviner…

```
[+] Detected sources (top 10)
- Twitter for iPhone    1406 (46%)
- Twitter for Android   1300 (43%)
- Twitter Web Client     267 (8%)
- Twitter for iPad        22 (0%)
- Instagram                2 (0%)
- Media Studio             1 (0%)
- Twitter Ads              1 (0%)
- Periscope                1 (0%)
```

Sources des tweets du compte de Donald Trump

Recommandations générales

Je vous recommande fortement de lire les conseils de sécurité Twitter du Grugq. En plus de ce guide, je vous conseille d'être prudents avec les fuseaux horaires et les langues que vous utilisez, et d'être également conscients que vos tweets peuvent être analysés comme un tout : ne tweetez pas toujours à la même heure si vous ne voulez pas que les gens devinent votre fuseau horaire. Bien sûr, ces principes sont valables seulement si vous souhaitez rester anonyme, ne les appliquez pas à votre compte principal (ce serait une perte de temps) !

Code source

J'ai publié mon script sur GitHub. C'est *open source* donc n'hésitez pas à contribuer