# Trafic non-humain sur Internet

## alerte aux faux internautes

Selon un rapport publié en décembre dernier*, le trafic robotisé représentait en 2013 plus de 60 % du trafic total sur l'Internet. Cette proportion est en augmentation de plus de 20 % par rapport à 2012 et rappelle un fait simple aux implications complexes : sur le web, il n'est pas toujours facile de faire la différence entre humains et robots.

L'augmentation rapide du trafic non-humain sur le réseau a des causes multiples dont toutes ne sont pas problématiques pour le marketing en ligne. Activité accrue des robots d'indexation, multiplication des outils d'optimisation du référencement naturel et des logiciels de veille digitale : ces agents non-humains sont toujours davantage présents sur le réseau. La plupart rend d'ailleurs des services tangibles aux internautes. Ces espions « bienveillants » permettent notamment aux moteurs de recherche des mises à jour toujours plus fréquentes, dans le but de fournir des résultats toujours plus pertinents et actuels. Mais le rapport paru l'hiver dernier tire aussi la sonnette d'alarme : la forte croissance du trafic non-humain est également due à des robots malveillants et notamment à des « imposteurs », des logiciels dont l'apparence informatique est celle d'un internaute conventionnel.

Une requête en ligne sur quatre est aujourd'hui provoquée par un logiciel singeant un internaute.

Pour un adserver ou un système d'évaluation de trafic, la manière la plus simple de filtrer le trafic non-humain est de retrancher de ses statistiques tout trafic généré par des agents signalés sur la liste des « Spiders & Robots » de l'IAB / ABCe. Cette liste, qui paraît depuis 2004, tient à jour un recensement des robots et autres « collecteurs », bienveillants ou malveillants, et est accessible à tous les acteurs de la chaîne digitale. Elle permet une filtration automatique de l'activité

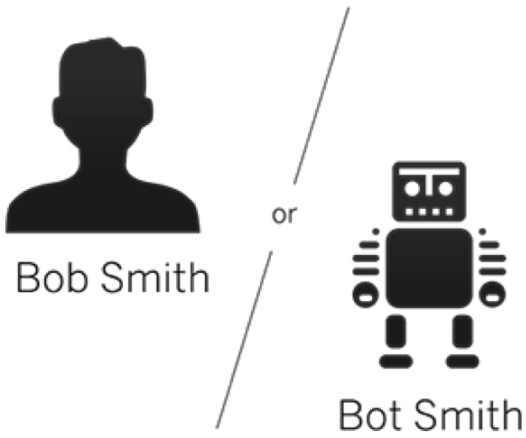
* paru sur le blog d'Incapsula, société active dans la sécurité digitale, http://www.incapsula.com/blog/bot-traffic-report-2013.html

robotique connue, que cela soit dans les statistiques de trafic d'un site ou dans celles d'une campagne publicitaire. Mais pour les robots « imposteurs », la simple filtration via cette liste est insuffisante.

Seule une analyse comportementale du trafic permet de distinguer des humains de tels logiciels. En clair, il s'agit de repérer dans la manière dont un navigateur se comporte sur le web des schémas suspects, qui laissent penser à une activité non-humaine ( cycles de consultations, de clics, récurrence des sites consultés, etc. ). On imagine ici la difficulté d'un tel exercice et les ressources « big data » nécessaires pour le mener à bien. Selon les différentes estimations disponibles, ce type de robot imposteur ou « impersonator » générerait aujourd'hui entre 20 % et 35 % du trafic global, ce qui signifie grosso modo qu'une requête en ligne sur quatre est aujourd'hui provoquée par un logiciel singeant un internaute, cela sans que des moyens sûrs d'identifier ce type de trafic existent.

L'IAB n'hésite plus aujourd'hui à parler d'une véritable « crise » du trafic frauduleux.

Mais qui est derrière ces faux internautes ? Et dans quel but sont-ils lancés sur la toile ?




www.whiteops.com

L'an dernier, le marché de la publicité en ligne de type display représentait aux Etats-Unis un investissement total d'environ 13 milliards de USD (source PwC). Un quart de cet investissement est aujourd'hui réalisé de manière automatisée via des places de marché en ligne, des bourses d'impressions publicitaires, les fameux Ad Exchanges. La caractéristique de ce type d'achat RTB / RTA ( Real Time Bidding / Advertising ) est de pouvoir acheter non plus un placement précis sur un site, mais des impressions sur un type d'internaute bien défini, au moment même ou cette impression est consommée et cela souvent au travers d'un système d'enchère. Dans ce principe de planification, c'est donc moins la source de trafic ( la page où l'impression est effectivement délivrée ) qui est en jeu que la pure performance publicitaire et son prix. Les indicateurs ROI ont ici le premier rôle et des algorithmes optimisent automatiquement les achats sur la base de ceux-ci. Le but est simple : obtenir un maximum d'impressions, de clics ou de conversions avec un minimum d'investissement.

Au niveau global, tout indique que la croissance de l'achat programmatique (RTA / RTB) coincide avec celle du trafic non-humain malveillant. Dans un environnement d'achat dématérialisé, où l'attention se porte beaucoup plus sur la performance publicitaire que sur les sources de trafic, on comprend aisément que la capacité à générer des impressions ou des clics en grand nombre puisse se révéler financièrement très attractive pour les producteurs d'un tel trafic.

La polémique enfle ainsi aux Etats-Unis depuis le début de l'année. Là-bas, l'IAB n'hésite plus aujourd'hui à parler d'une véritable « crise » du trafic frauduleux. Le Bureau a d'ailleurs mis sur pied une nouvelle task force dont l'objectif est de mieux comprendre et faire connaître le problème du « non-intentional traffic » au sein de l'industrie (voir : www.iab.net / fraud).

Les techniques de fraudes sont en effet de plus en plus sophistiquées, leur évolution culminant dans le « botnet ». Plus qu'un simple programme informatique générant des requêtes, le botnet est un réseau constitué d'ordinateurs privés dont une partie de l'activité est pilotée à distance par un logiciel propagé via un virus informatique (aux Etats-Unis, ces ordinateurs « infectés » représenteraient 15 % du parc informatique privé connecté en haut débit). Il est alors possible de faire surfer massivement, et à l'insu de leurs propriétaires, des terminaux réels avec des caractéristiques et des cookies eux aussi bien réels, sur des sites désignés par le pilote du botnet. Ces réseaux « zombies » qui permettent classiquement aux pirates informatiques de mettre sur pied des attaques de type DDoS sont ainsi toujours plus utilisés aujourd'hui pour mener de telles fraudes publicitaires.

## Whiteops, spider.io : la résistance contre les robots s'organise.

La société spécialisée américaine Whiteops estime le chiffre d'affaires actuel de la fraude publicitaire en ligne à plusieurs milliards de dollars. Son CEO, Michael Tiffany, expliquait lors de l'une récente conférence de l'IAB avec un exemple très simple pourquoi celle-ci est en vogue chez les hackers : sur le marché noir, un numéro de carte de crédit volé vaut 25 cents, alors qu'un millier d'impressions publicitaires générées via un logiciel rapporte facilement dix fois davantage tout en impliquant nettement moins de risques sur le plan juridique.

Un autre expert reconnu dans ce domaine est Douglas de Jager, fondateur de spider.io, une société rachetée en février dernier par Google. Sa société est notamment à l'origine de la découverte du botnet « Chameleon » au printemps 2013. Ce réseau générait, au moment de sa découverte en mars 2013, environ 30 % de l'inventaire publicitaire disponible sur les principaux Ad Exchanges américains et a également touché les sites d'éditeurs importants. Les campagnes des plus grandes marques ont alors été piégées par le botnet - Amex, Ford, BMW ou encore McDonald's. Malgré l'ampleur de la fraude, aucun coupable n'a pu être jusqu'à aujourd'hui identifié derrière « Chameleon ».

Face à l'amplification du phénomène, des solutions sont développées pour contrer ou tout au moins réduire les risques d'erreurs et de fraudes dans la mesure du trafic en ligne. Mais les fraudeurs semblent avoir aujourd'hui une bonne longueur d'avance, cela en raison d'un marché publicitaire lent à réagir, craignant de mettre en péril la réputation de mesurabilité « absolue » du digital et la formidable croissance des investissements liés au nouveaux médias.

## Comment réagir face à ce phénomène ?

Comme dit précédemment, le recours à la liste « Spider & Bot List » de l'IAB est la première étape pour tous les acteurs de la chaîne numérique. Notre expérience montre que certains adservers utilisés par des éditeurs d'importance en Suisse peuvent montrer des faiblesses déjà à ce stade, dans la mesure où la mise à jour de cette liste n'est pas réalisée aussi souvent que nécessaire. D'autre part, certains Ad Exchanges proposent encore aujourd'hui des inventaires comprenant une importante proportion d'agents non-humains présents sur cette liste. Il convient ici d'interroger ses partenaires technologiques et commerciaux pour s'assurer que cette filtration minimale soit activée.

Les technologies de filtrage comportemental sont elles aussi en constante évolution et sont également une étape indispensable pour tout acteur responsable de l'écosystème Online Display. Difficile d'évaluer ici la situation, dans la mesure où ces techniques de filtrages sont évidemment gardées secrètes. Un annonceur ou une agence est malgré tout en droit de demander à ses partenaires de présenter leur politique en matière de filtration comportementale et les éléments qui la composent.


Adweek, octobre 2013

En termes d'adserving, l'évaluation de la visibilité de moyens publicitaires est enfin à notre sens un élément primordial pour la surveillance d'éventuelles irrégularités dans la diffusion d'une campagne.

En 4 points, voici à notre avis les mesure et attitudes à adopter dans le contexte de l'augmentation du trafic non-humain et du risque de fraude sur l'Internet :

1. Adserving en partie tierce : pour une campagne display, le recours à un adserver indépendant du vendeur de l'espace est l'étape élémentaire pemettant de découvrir un éventuel problème de fraude. Sans cela, l'annonceur reste extrêmement vulnérable à une éventuelle faiblesse de l'éditeur du site en matière de filtration du trafic non-humain, ou à tout autre type de fraude.

2. Capacités avancées de filtration du trafic robotisé :
   La certification (par exemple MRC) des métriques disponibles dans une solution d'adserving est un élément à valoriser davantage. Mais la capacité d'une solution d'adserving à détecter le trafic non-humain, et notamment les fraude de type « botnet », apparaît comme un élément de plus en plus important dans la sélection et le choix d'une telle solution.

3. Vigilance accrue sur les activités RTB / RTA : lors d'un achat programmatique en temps réel, les sources d'inventaires sélectionnées doivent pouvoir être le plus transparentes possibles et la visibilité des moyens publicitaires pouvoir y être évaluée systématiquement. Pour des annonceurs avec un volume d'achat important sur les Ad Exchanges, l'évaluation d'une solution de détection anti-fraude

en partie tierce nous paraît aujourd'hui se justifier ( par exemple Peer 39 de mediamind / Sizmek ).

4. Critères d'optimisation automatique à définir avec précaution : tout élément automatisé de la chaîne de planification doit être considéré avec une grande prudence. On pense notamment ici à des algorithmes qui peuvent avoir des effets pervers dans le cas de fraude - par exemple, avec un algorithme d'optimisation de la diffusion en fonction du taux de clic dans le contexte d'un site accueillant un important trafic robotisé générateur de clics. Les campagnes de type RTB / RTA devraient d'ailleurs autant que possible éviter une optimisation automatique sur des métriques facilement manipulables comme les clics ou les interactions et fixer des objectifs liés à des conversions « réelles ». L'important ici est de ne pas prêter exclusivement attention au rapport prix / performance mais de s'interroger de manière critique sur la nature de la performance réellement fournie.

Sur le web davantage encore que sur d'autres canaux, compter n'est pas mesurer. Une visite, une impression, un clic ou une conversion ne sont pas des « faits » mais des indicateurs résultants d'un filtrage et d'une interprétation de données brutes. Face à la multiplication d'acteurs malveillants, l'industrie publicitaire a tout intérêt à s'en souvenir et à mettre en place les mesures qui s'imposent pour sauvegarder sa crédibilité.

Matthieu Robert – Online Director
mediatonic SA