

Attention ! sites

Fabrice Prigent

RESPONSABLE RÉSEAU

CENTRE DE RESSOURCES INFORMATIQUES

UNIVERSITÉ DE TOULOUSE-I SCIENCES SOCIALES

Une même liste noire préside au filtrage d'Internet dans la plupart des établissements français, équipés d'un serveur SLIS, Pingoo, Eole ou Linuxedu : c'est celle de l'université de Toulouse-I, mise à jour en permanence. Comment est-elle conçue ? Dans quel état d'esprit ? Pour quels usages ?

La liste noire de l'université de Toulouse-I est une collection d'adresses web qui ont été classées en diverses catégories afin, généralement, d'en interdire l'accès. Elle est utilisée majoritairement par les écoles, collèges et lycées, par l'intermédiaire de divers serveurs web. Elle se concentre essentiellement sur la catégorie « adulte », qui est la plus importante en volume.

Comment est constituée la liste noire ?

Un programme informatique, ayant pour nom Pornperl, a été conçu en 2003 pour rechercher les sites de la catégorie « adulte ». Il a d'abord été mis au point à partir de plus de 100 000 sites internet dans diverses langues, afin d'apprendre à distinguer les sites ordinaires des sites « adultes ». Depuis, chaque jour, il explore Internet et, après évaluation, rapporte entre 100 et 300 adresses supplémentaires, ce qui monte la liste noire à près d'un million d'adresses web actuellement.

La base est également enrichie par la collaboration d'une vingtaine de personnes de par le monde, qui ajoutent, dans les diverses catégories, de nouvelles adresses.



sensibles et liste noire

Qu'est-ce qu'une page « adulte » ?

Lors de son apprentissage, Pornperl a pris chaque page qui lui a été présentée et l'a décomposée en binômes de mots. Chaque couple recevait un point négatif quand la page dont il provenait était pornographique, et un point positif quand elle ne l'était pas. Après l'apprentissage, la proportion de points négatifs et de points positifs donne une note d'« adulticité » du binôme. Celui-ci est conservé avec sa note dans une base de données.

Une fois l'apprentissage effectué, toute nouvelle page à évaluer est décomposée en couple de mots, dont on cherche ensuite la note. La page est alors affectée d'une note correspondant à une sorte de moyenne de la note de chacun des binômes qui la composent.

Le seul *a priori* est donc dans le fait d'avoir, au départ, choisi les pages à caractère pornographique. Le reste n'est plus qu'une histoire de statistiques.

À quoi sert une liste noire ?

La fonction d'une liste noire est d'empêcher l'accès à des ressources, dans notre cas des sites web. Mais la véritable question est : pour quelle raison me servirais-je d'une liste noire ? Là, les choses se compliquent, car si l'outil suggère un usage, ce sont les objectifs qui, au final, vont lui donner une réalité. Voyons quelques-uns des usages les plus courants :

– empêcher l'utilisation abusive de ressources, qu'il s'agisse des ordinateurs d'un CDI, d'une bibliothèque ou bien de la bande passante de l'établissement connecté à Internet. C'était d'ailleurs l'objectif initial qui a amené à la création de la liste noire ;

– respecter la loi : nombre de textes de lois, règlements, contrats ou simples jurisprudences peuvent nécessiter le recours à un tel outil. On notera en particulier l'article 227-24 du code pénal sur la protection des mineurs, ou la directive rectorale du 6 janvier 2004 ;

– améliorer la productivité ou, plus précisément, éviter les sources de distraction telles que sites de presse, de jeux, de forums, etc. ;

– pour le respect des personnes et de la dignité humaine, notamment en ce qui concerne les sites xénophobes, ou ceux présentant des situations dégradantes pour des camarades de classe, voire des enseignants, tels que certains blogs collégiens ou lycéens ;

– pour des raisons morales. Personne jusqu'à présent ne m'a annoncé un tel objectif, mais pourquoi pas ?

– pour des raisons de sécurité : trois aspects différents peuvent être couverts sur ce thème : la sécurité physique (les sites expliquant la conception de bombes, de produits chimiques dangereux), la sécurité financière (l'ensemble des sites de hameçonnage, aussi appelés sites de *phishing*, qui se font passer pour des sites bancaires et permettent de voler les codes) et la sécurité informatique (les sites qui sont connus pour favoriser la propagation des virus informatiques). On notera d'ailleurs que des sites adultes russes avaient acheté un programme informatique, de type cheval de Troie, spécifiquement conçu pour infecter les ordinateurs de leurs visiteurs.

Ces dangers ne sont-ils pas exagérés ?

La nature humaine est ainsi faite que les deux moteurs les plus efficaces sont le sexe et l'argent. Le premier est un levier tel qu'il est couramment admis que c'est le marché du sexe qui a fait gagner les cassettes VHS contre le BetaMax, pourtant de meilleure facture, parce que le VHS avait réussi à faire émerger le minitel, grâce au minitel rose. C'est aussi ce qui explique les communiqués du consortium DVD Blu-ray (un des deux formats haute définition du DVD) concernant l'adhésion d'un célèbre fournisseur de films X à son format, ceci afin de contrer les rumeurs insistantes sur l'adhésion quasi totale des autres fournisseurs au format concurrent.

Or, là où il y a un besoin, il y a un fournisseur qui s'intéresse au profit. Et plus le besoin est « tabou », plus l'argent est facile à obtenir. Ceci explique que les fournisseurs les moins scrupuleux soient tentés, pour attirer le chaland, d'aller jusqu'à « détourner » les adresses de sites web

Il a d'abord été mis au point à partir de plus de 100 000 sites internet dans diverses langues, afin d'apprendre à distinguer les sites ordinaires des sites « adultes »...

... Depuis, chaque jour, il explore Internet et, après évaluation, rapporte entre 100 et 300 adresses supplémentaires, ce qui monte la liste noire à près d'un million d'adresses web actuellement.

ordinaires, tels qu'un site de prévention routière, de firme de matériel aéronautique ou d'université. Ou, de manière moins illégale, d'acheter des noms de domaines approchants (par exemple, les domaines france2.com et france3.com, qui n'ont aucun rapport avec nos chaînes nationales).

D'autres cherchent aussi l'argent directement en volant des identifiants grâce à la réalisation de sites imitant à la perfection des sites bancaires. De manière plus détournée, l'utilisation de sites familiers (FAI, sites d'enchères, etc.) demandant de « re-saisir » des renseignements personnels, comme le numéro de carte bancaire, est une pratique très courante.

D'autres éléments peuvent motiver certains individus : le fanatisme, la haine, le jeu. Mais leur impact sur « l'Internet de tous les jours » est à un niveau différent, plus stratégique.

L'économie illégale sur Internet

C'est un modèle très simple, et très efficace : des vendeurs de produits pharmaceutiques frelatés, de rachat de crédits à des taux usuriers, d'accès payants à des sites pornographiques, veulent faire de la publicité à peu de frais. Ces vendeurs s'adressent à des spammeurs. Ces derniers proposent de diffuser leur publicité auprès de centaines de millions d'utilisateurs pour un coût dérisoire (le taux communément admis est qu'une réponse sur mille suffit au vendeur pour rentrer dans ses fonds).

Ces spammeurs ont besoin de milliers d'ordinateurs pour envoyer des spams. Pour cela, ils s'adressent à des pirates informatiques.

Les pirates créent des virus informatiques qu'ils diffusent par mail, ou, de plus en plus, à l'aide de sites web qui infectent les ordinateurs des visiteurs. Ceux-ci passent alors sous le contrôle du pirate, qui va pouvoir vendre ou louer ces dizaines de milliers de machines « compromises ».

Quelques chiffres : ce marché représente plusieurs milliards de dollars par an ; une faille de sécurité inconnue peut se vendre 80 000 euros ; la location de 10 000 ordinateurs (pour l'envoi de spams) se monnaie 100 euros de l'heure. Un bon pirate informatique peut toucher 15 000 euros par mois.

Quels sont les inconvénients du filtrage ?

L'utilisation de logiciels de filtrage (basés ou non sur les listes noires) comporte deux inconvénients notoires :

- le risque de blocage d'un site normal. C'est

ce que l'on appelle un « faux positif » ;

- le non-blocage d'un site interdit. C'est ce que l'on appelle un « faux négatif ».

Chacun des deux cas pose problème mais, selon le public visé et le contexte, les conséquences sont variables. Pour de jeunes enfants, le second inconvénient est le plus critique : mieux vaut bloquer un site correct, plutôt que de laisser passer un site malsain. On peut avoir recours, dans ce cas, à des listes blanches, qui contiennent les seuls sites autorisés.

A contrario, à moins de vouloir une rébellion, voire pire, la mise en place d'un contournement généralisé du filtrage (cela s'est déjà vu), il vaut parfois mieux laisser passer des sites douteux que de bloquer un site médical, si l'on s'adresse à des adolescents ou à des adultes. De même, le ralentissement excessif des communications dû à une analyse trop poussée peut rendre impraticable la navigation sur Internet.

D'autres inconvénients, plus diffus, émergent avec le temps en faisant apparaître une trop grande confiance dans l'outil, conduisant à laisser des enfants seuls devant Internet, ou un désir de rébellion pour « chercher la faille ». Cela ne se manifeste généralement qu'avec les adolescents.

Le dispositif peut-il être contourné ?

Le contournement du filtrage n'est pas impossible mais il dépend de la manière dont celui-ci est installé. Les systèmes qui se basent sur des ordinateurs dédiés (SLIS, Eole, etc.) sont les plus difficiles à déjouer. Des méthodes compliquées et peu pratiques existent mais elles aboutissent très rapidement à un blocage assez frustrant (une page avec uniquement du texte...).

Les logiciels de filtrage qui sont directement sur les ordinateurs sont plus facilement déverrouillables, car directement accessibles. D'autant plus que l'imagination des grands frères ou des amis peut donner un sérieux coup de pouce à un collégien motivé. Mais cela n'est tout de même pas à la portée de tout le monde.

Et si je débranchais simplement l'ordinateur ?

Certaines personnes finissent par hésiter à laisser l'accès libre à l'ordinateur. Mais refuser la réalité n'est pas la solution pour s'en protéger. C'est l'accompagnement et l'éducation, sur les chemins de l'information, en sécurité, qui permettent de découvrir les plaisirs et les dangers de cet espace de communication. Plus que jamais, l'étude du verbe, de l'image, et de leur mise en perspective révèle son importance. ●