Lutter contre les Spams Image

Des techniques d'envoi de spams en perpétuel renouvellement

Les spammeurs développent des méthodes de plus en plus sophistiquées pour contourner les filtres antispam. En général, la technique utilisée consiste à envoyer des "vagues" d'e-mails dans lesquelles chaque message est unique et légèrement différent de ceux envoyés précédemment. Le succès de chaque vague est ensuite analysé par le spammeur et les résultats "positifs" obtenus deviennent des "caractéristiques" intrinsèques pour la vague de spam suivante.

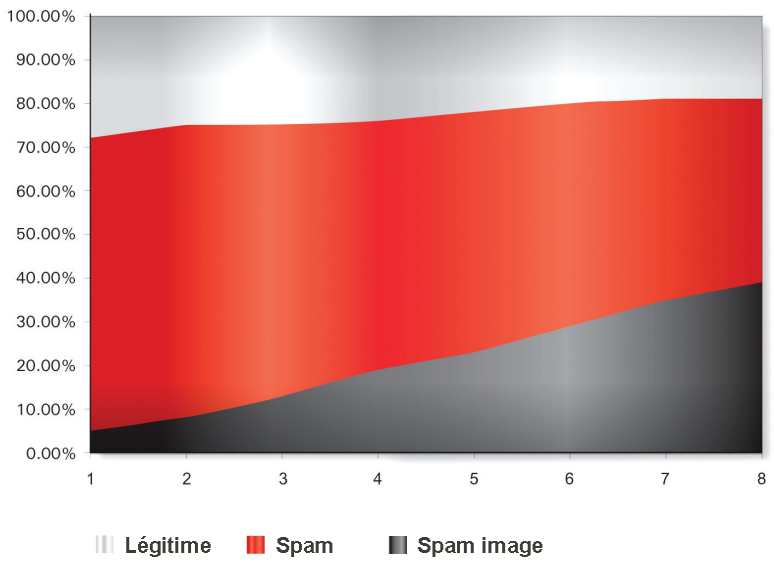
Les nouvelles méthodes pour détecter les vagues de spams, comme l'extraction de leurs caractéristiques et leur classification sous forme de signatures spam, sont en phase finale de développement. Des recherches sont également menées afin de trouver des méthodes permettant d'anticiper les évolutions des spams.

La majorité des méthodes de filtrage utilisées par BitDefender permet désormais de gérer les petites variations de contenu utilisées dans les vagues de spams. Cependant, la quantité de spams images a crû énormément depuis 2006. Des e-mails simples avec des images apparemment identiques (mais qui sont en réalité uniques, à en juger par leurs différences informatiques) ont commencé à "polluer" nos boîtes de réception de manière préoccupante.
A l'époque où les premières techniques de lutte contre le spam image ont été mises en place, la méthode la plus efficace pour détecter un spam image consistait à établir des signatures basées sur les métadonnées de l'image. Néanmoins, le laboratoire antispam BitDefender a détecté dans le même temps de nombreux spams en circulation utilisant de toutes nouvelles techniques de spam image, élaborées spécialement pour contourner les filtres antispam et il est donc devenu nécessaire de mettre au point une nouvelle technologie permettant de contrer ces nouveaux développements.

La première approche

En 2006, le "spam image" a représenté près de 10% du total des spams en circulation. Ces séries de messages consistaient principalement en 5-6 spams images contenant de légères modifications. Les spammeurs, ayant rapidement constaté que la plupart des solutions antispams actuelles étaient inefficaces face à ce nouveau phénomène, se sont donc empressés d'exploiter cette niche. Le "spam image" a ensuite représenté 30 à 40% du total des spams en circulation, avec des changements de "bruit" dans les images intervenant de manière aléatoire, quasiment pour chaque image envoyée. Les taux de détection des antispams ont ainsi chuté, passant de plus de 97% à presque 65-75%. Les spams images contiennent généralement des images de Viagra, du matériel informatique, des images pornographiques, ou simplement un message spam classique (du texte et une URL), mais écrit dans une image de mauvaise qualité. Réaliser une analyse du contenu de ces emails quelle qu'en soit la forme, impliquerait d'analyser les images avec un outil de reconnaissance optique de caractère (OCR). Cependant, la plupart des moteurs OCR sont très consommateurs en ressources machine et leur précision sur de telles images laisse beaucoup à désirer.

Evolution des Spam Image sur les 8 premiers mois de 2006 :



bitdefender
secure your every bit

L'approche de BitDefender

Pour obtenir une détection plus fiable, BitDefender propose une alternative à la technique OCR se basant sur des méthodes d'apprentissage qui utilisent l'expérience aquise afin de détecter les caractéristiques communes aux images elle-mêmes. Cette alternative repose sur l'utilisation de deux techniques, l'extraction et la comparaison d'histogrammes*, techniques qui se sont avérées efficaces dans des applications impliquant des traitements d'image. Elles sont généralement utilisées dans le cadre d'extraction d'images en fonction de leur contenu (par exemple, l'extraction d'images contenant des palmiers parmi plusieurs photos), avec un taux de "faux-positifs" assez élevé. A l'usage, cette solution s'est avérée relativement problématique en raison de son fort taux de "faux positifs", qui signifiait potentiellement la perte d'un certain nombre d'emails pour l'utilisateur.

Une nouvelle technologie permet d'obtenir un faible taux de faux positifs : l'algorithme SID (Spam Image Distance) répertorie les images sur la base de leur ressemblance en terme de nombre de points et de similitudes dans les couleurs plutôt que sur la base de leur ressemblance au niveau de la forme. L'algorithme SID permet de distinguer les images les unes des autres avec une nouvelle approche. Par exemple, bien que toutes les images disponibles sur des pages imprimées se ressemblent plus ou moins, en étant soit "blanches" soit non "blanches", avec des quantités de gris plus ou moins sombres, une page d'encyclopédie ne ressemble pas à une page de publicité dans un magazine en raison de l'écart qui existe dans la quantité de gris utilisé.

La technique SID est utilisée pour comparer les images et évaluer la "distance" qui les sépare, ce qui revient essentiellement à trouver ce qui les diférencie. La distance trouvée grâce à l'algorithme SID est utilisée pour comparer les images qui sont déjà incluses dans la base de données de spam avec celles qui pourraient en être. Si l'analyse de l'image renvoie un score inférieur à un certain seuil, l'image est alors classée dans la base de spams images de BitDefender. C'est pourquoi le SID est une technique de premier ordre pour permettre l'indexation des spam images qui sont des variantes d'autres spams images plus anciens. Ces nouvelles techniques ont prouvé leur efficacité sur des images assez "propres" mais le problème des images ayant subi des altérations (ajout de bruit) subsiste. Heureusement, les techniques d'altération utilisées par les spammeurs sont relativement bien connues et l'arsenal de contres mesures disponible est pareillement bien fourni. Par exemple, les spammeurs vont découper une image en sous images et les placer les unes à côté des autres dans une page HTML afin de reconstruire l'image de départ. Cette technique peut être contournée en reliant les différents histogrammes obtenus à partir des sous images et en recréant celui de l'image originale afin de l'analyser avec l'algorithme SID.

Technique d'ajout de "bruit" :

- Ajout aléatoire de pixels dans les images.
- GIF animés incluant des pages corrompues.
- Utilisation de couleurs similaires dans le texte et dans l'image.
- Une longue ligne à la fin de l'image (comme une bordure) avec des parties manquantes de manière aléatoire.
- Eclatement d'une image en plusieurs sous images recomposées dans un fichier HTML.
- Envoi de la même image mais avec des tailles différentes.
- Corruption d'image : par insertion d'images ayant un contenu légitime, comme des logos de société.
- Envoi d'images légitimes légèrement corrompues pour brouiller le filtrage.
- Envoi d'images légitimes avec un contenu proche d'un spam (par ex. des images relatives au crédit issues de vraies sociétés de crédit).

* Un histogramme peut être défini comme une liste mettant en relief la relative prépondérance de certaines couleurs dans une image, il indique de quelle couleur il s'agit et combien de pixels sont utilisés pour chaque couleur.

Taux de détection

Cette technologie, pour laquelle une demande de brevet est en cours, offre un taux de détection de 98,7% au sein du corpus BitDefender de spams images (quelques millions d'échantillons extraits de vrai spams). 1,23% de ces images sont déformées, ce qui signifie que leurs histogrammes ne peuvent pas être extraits mais qu'elles ne peuvent pas non plus être affichées. Les 0,07% restant concernent les faux positifs. Si on enlève les images déformées du corpus, le taux de détection monte quasiment jusqu'à 100%. Avec des résultats aussi prometteurs, l'algorithme SID constitue un atout précieux pour enrichir l'arsenal des solutions antispam modernes; de plus, les progrès réalisés pour réduire le bruit sur les images devraient permettre d'améliorer encore le potentiel de cet outil déjà particulièrement efficace.

La gamme BitDefender®
est éditée dans les pays francophones
en exclusivité par Editions Profil.

ÉDITIONS PROFIL
49, rue de la Vanne - 92120 Montrouge
Tél. (+33) 1 47 35 72 73
Fax (+33) 1 47 35 07 09
E-mail : bitdefender@editions-profil.eu
www.editions-profil.eu

bitdefender
secure your every bit