>> Jorge Dávila

La nueva plaga se llama *spam*

El nivel de tráfico de correo electrónico en Internet debido al correo no solicitado (*spam*) ha aumentado exponencialmente en los últimos meses. Lo que era una muestra de desfachatez por parte de algunos, se está convirtiendo en una plaga para todos. En este contexto, las empresas de seguridad están proporcionando herramientas para defenderse, pero no todas ellas son igualmente "buenas" y habrá que utilizarlas con cuidado. Sin duda, se impone tanto la investigación propia como la colectiva para poder atajar ataques promovidos por comunidades que sí investigan cómo salirse con la suya.

Últimamente se ha puesto de moda un problema que no es realmente nuevo y que se conoce como SPAM. En los últimos meses todos hemos podido comprobar en nuestras carnes cómo ha aumentado el número de mensajes en nuestras cuentas de correo electrónico. Esos mensajes, o bien nos intentaban infectar con el último virus, gusano o similar, o bien nos querían alargar alguna parte de nuestra humilde fisonomía, o nos ofrecían la ganga más increíble a precios asequibles, o los fármacos (reales y milagrosos) que no podemos conseguir sin pasar por la consulta del médico, o algún negocio turbio para sacar a circulación alguna inconfesable fortuna de algún presunto líder –siempre caído en desgracia–, de Nigeria. A esto se le denomina *spam*, a la "recepción de correo electrónico no solicitado". La definición de *spam* no es del todo nítida ya que, con la anterior, nos cargaríamos la posibilidad de hacer publicidad a través del correo electrónico, lo cual quizás sea una solución.

Antes, para evitar esa abrumadora molestia, recurríamos a borrar todo aquello que no viniese de un remitente conocido, pero eso ya no sirve porque los mensajes de *spam* desde hace tiempo suplantan direcciones que sí pueden parecernos interesantes. Aunque borremos los mensajes, éstos han circulado por la red y han consumido recursos humanos y materiales que no están montados para facilitar la publicidad de los *spammers*. Así pues, la preocupación de las empresas ya es grande y la alarma parece haberse establecido ahora en este tema. Como respuesta a ello, algunas compañías de seguridad, cómodamente instaladas en la lucha contra los virus, han decidido incluir soluciones anti-spam en sus productos.

Si el spam no desaparece o disminuye a límites anteriores a la "comercialización" de Internet, lo que sí disminuirá hasta desaparecer será la utilidad social del correo electrónico.

Son varias las técnicas que se han propuesto y están en marcha para intentar resolver este fenómeno, pero el problema esencial de cualquier sistema de correo electrónico que pretenda rechazar el *spam* es que, en su celo, el sistema también rechaza correos legítimos. A estos fallos del sistema se les denomina "falsos positivos" en la terminología de los filtros.

Técnicas utilizadas por la industria beligerante en esto del anti-spam son, a grandes rasgos, los que vienen a continuación:

Listas Blancas y Negras[1]: es la aproximación más simple y clásica al problema del *spam*. Consiste en mantener actualizadas listas de dominios o usuarios legítimos (blanca) y de *spammers* (negra). Son medidas muy eficaces en cuanto a su tasa nula de "falsos positivos", ya que los nombres apuntados en la lista blanca prevalecen sobre cualquier otra consideración. Ambas listas requieren un mantenimiento muy detallado y actualizado por lo que esta posible solución requiere un gran esfuerzo humano.

Análisis de cabeceras: este es el método más extendido entre la mayoría de presuntas herramientas anti-spam. Consiste en analizar las cabeceras de los mensajes recibidos e intentan probar su "validez" y la existencia de la dirección electrónica del remitente.

Análisis de contenido: esta aproximación se basa en analizar lo que viene dentro del mensaje a la caza de determinadas secuencias de caracteres o palabras que hayan sido definidas por el sistema como propias del *spam* y que sean impropias del correo comercial legítimo.

Análisis contextual: esta es una versión más sofisticada que la anterior y busca patrones sintácticos propios de los anuncios y la publicidad.

Heurística: las técnicas de aprendizaje automático, salidas de la Inteligencia Artificial, pretenden hoy establecerse en el frente anti-spam. Los filtros heurísticos analizan los mensajes declarados como *spam* y buscan características que los identifiquen como tales frente a otros que no lo son; de este modo pretenden "aprender" cuál es la diferencia entre unos y otros. Al igual que en el caso de la lucha anti-virus, los filtros heurísticos persiguen definir una "firma" que sea propia de lo que llamamos *spam* y luego utilizarla para aceptar o rechazar mensajes. En este grupo encontramos los muy publicitados filtros basados en técnicas estadísticas como es el Análisis Bayesiano. En este análisis se compara un conjunto equilibrado de mensajes declarados como *spam* y otros tantos legítimos. La "digestión" algorítmica de esta muestra conduce a ciertos valores umbral, mediante los cuales "separar" el flujo de mensajes entrante[2].

Filtros de reto/respuesta: la idea es no entregar un mensaje recibido en el servidor de correo hasta que el remitente demuestre que es un ser humano. Una vez recibido un mensaje, se genera otro dirigido al remitente en el que se le pide que mire a una figura y teclee el texto que aparece en ella[3]. AOL y Yahoo utilizan esta técnica para evitar que los *spammers* utilicen sus servidores de cuentas de correo como plataforma para sus actividades. MailFrontier y Mailblocks también emplean esta aproximación.

[1] Dentro de este grupo se incluye la iniciativa experimental PUAS de la RedIris que, por decisión política de su dirección, pronto se suspenderá; dejando así a la comunidad académica sin defensa alguna ante el *spam*, y a todos los demás usuarios de la red sin los resultados de la única iniciativa seria dentro del sector público español.

[2] Muchos programas anti-spam utilizan ya la heurística. Algunos ejemplos son: Sopho's PureMessage, Mirapoint's MessageDirector, Lyris MailShield Server y el programa de código abierto SpamAssasin.

[3] Se supone que, por el momento, las máquinas no saben "leer" un texto en formato imagen al que se le añaden distorsiones aleatorias y un importante nivel del ruido. Consiste en colocarse por delante de la tecnología de OCRs (*Optical Character Recognition*).

Cuentas cebo (*Honeypots*): esta es una técnica novedosa que utiliza cuentas de correo indistinguibles de todas las demás, pero que detrás no tienen a un usuario real. Estas cuentas se publicitan del mismo modo que las demás, por lo que, se supone, que el *spammer* las incluirá en sus listas; el correo que llegue a estas direcciones se considera *spam* y no se distribuye a las demás cuentas que sí tienen un usuario detrás. La compañía Brightmail dispone de cientos de miles de direcciones de correo distribuidas estratégicamente en toda la red, de modo que los correos que llegan a ellas son automáticamente considerados como *spam*. Con esa información la compañía crea reglas de filtrado que inmediatamente envía a sus clientes para eliminar esos mensajes de sus servidores de correo.

INCONVENIENTES DE LA TECNOLOGÍA DISPONIBLE

Las listas son una buena solución si sólo se quiere penar a los culpables demostrados de *spam*, pero hay varios problemas. Para empezar, la dirección (de correo o IP) desde donde se hacen los envíos podría cambiar de una vez a otra, por lo que las listas negras serían inmensas y la actualización de la lista sólo puede ser reactiva. También puede darse el caso de que una empresa o usuario que aparece en la lista blanca sea víctima de un ataque y comience a emitir *spam* por el efecto de una infección de un gusano *spammer*, por lo que todos sus envíos superarían el filtro y el sistema antispam resultaría burlado. Las listas blancas indican a los promotores del *spam* cuáles son los nodos que deben asaltar y utilizar para sus propósitos.

El análisis de cabeceras está obsoleto porque, desde hace ya tiempo, los *spammers* modifican los contenidos de la cabeceras de sus mensajes para suplantar direcciones válidas que no son suyas, y esto invalida cualquier conclusión sacada de las cabeceras. Un ejemplo de ello lo tenemos en el *spam* que persigue la infección con virus y gusanos a través de correo electrónico; los detectores de antivirus, al detectarlos, se han convertido en *spammers* involuntarios ya que responden con una alerta de infección a direcciones de correo desde las que nunca se enviaron dichos mensajes.

El análisis contextual muchas veces termina desbocándose, ya que las fronteras entre los leguajes publicitarios y los normales en muchas empresas es realmente difuso. Son sistemas con alto éxito en el bloqueo del *spam*, pero pueden llegar a tener un número nada deseable y no permisible de falsos positivos.

La industria de la seguridad debe afrontar este problema con visión de futuro y no simplemente echando mano de las soluciones materializadas en forma de pequeñas empresas; la investigación propia y colectiva es esencial para poder atajar ataques promovidos por comunidades que sí investigan cómo salirse con la suya.

En cuanto a los métodos heurísticos, hay que tener en cuenta que el éxito de cualquier método de aprendizaje automático depende, entre otras cosas, de lo completa y representativa que sea la muestra utilizada para realizar el análisis y cuán diferentes sean los mensajes de un tipo y otro; cosa nada fácil que requiere mucha experiencia por parte de los que han de confeccionar la muestra.

Además de esto, debemos tener cuidado con no proponer una solución que se convierta en un peligro aún mayor. El análisis del contenido de los mensajes de correo electrónico puede terminar siendo un delito contra la intimidad de las personas. Si un sistema automático puede detectar mensajes "no deseados" y genera una alerta de algún tipo, lo que está haciendo es "leer" el mensaje y eso podría vulnerar el derecho personal al secreto de las comunicaciones.

Los curiosos filtros basados en protocolos de reto y respuesta son un modo muy primitivo de "autenticar" al supuesto remitente, y requieren la participación activa de éste en la respuesta a un segundo mensaje, lo cual puede llegar a ser un verdadero tostón. La autenticación, en serio, se conseguiría haciendo que los mensajes de correo electrónicos fuesen firmados digitalmente pero para ello necesitamos una galaxia de PKIs que se reconociesen entre sí y que la confianza hubiese colonizado Internet, lo cual es una quimera de las de verdad.

La técnica de cuentas señuelo es clara, sencilla y puede funcionar siempre y cuando la cualidad de "cuenta señuelo" quede en secreto. El éxito de este proceder depende de la capacidad que tenga la comunidad *spammer* de confeccionar listas negras de direcciones trampa. Este enfoque daría lugar a una curiosa batalla incruenta entre empresas proveedoras de datos provinentes es estos cepos y los generadores de listas de *spam*; unos tendrían que cambiar al azar la ubicación e identidad de sus trampas y otros tendrían que descubrirlas.

NO TODO ESTÁ PERDIDO

Queda otra posibilidad y es la de que los ISPs se involucren de verdad en esta guerra. El correo electrónico no se difunde por arte de magia, sino a través de la comunidad de ISPs, por lo que ellos podrían **contabilizar** el número de mensajes que está enviando cada uno de sus clientes. Para el usuario común lo normal es no mandar cientos de miles de mensajes al día o por hora, por lo que cualquier desviación de la "normalidad" en el envío de correo puede ser indicio de actividades de *spam* y habría que investigarlo. Para evitar que el ISP tenga que analizar el contenido de un correo que no va dirigido a él (lo cual sería un delito) lo que puede hacer es "lentificar" (técnica de los "pozos de brea") el re-envío de correos según la desviación de la normalidad aumenta. De este escenario habría que sacar a todas aquellas empresas que distribuyen sus servicios a través de envíos masivos de correo (listas de distribución, grandes entidades, etc.) y, a cambio de ese tratamiento especial, se comprometerían legal y responsablemente a no practicar ni permitir prácticas de *spam* desde sus equipos.

Hemos visto que los administradores de correo tienen un amplio rango de herramientas a su disposición y que hay nuevas ideas que investigar y poner a prueba en escenarios reales. La industria de la seguridad debe afrontar este problema con visión de futuro y no simplemente echando mano de las soluciones materializadas en forma de pequeñas empresas; la investigación propia y colectiva es esencial para poder atajar ataques promovidos por comunidades que sí investigan cómo salirse con la suya.

La historia reciente demuestra que el crecimiento del *spam* es muy importante, tanto en volumen como en variedad, por lo que habremos de pensar que no se han tomado medidas técnicas y, sobre todo, políticas, legales y sociales suficientes para terminar con esta lacra. Si el *spam* no desaparece o disminuye a límites anteriores a la "comercialización" de Internet, lo que sí disminuirá hasta desaparecer será la utilidad social del correo electrónico. ■

Jorge Dávila Muro
Director
Laboratorio de Criptografía
LSSI - Facultad de Informática - UPM
jdavila@fi.upm.es