Luchando contra la publicidad no deseada (o Spam) en el correo

by Katja y Guido Socher
<katja/at/linuxfocus.org
guido/at/linuxfocus.org>

About the authors:

Katja es la editora de la versión en alemán de LinuxFocus. A ella le gustan Tux, el cine, la fotografía y el mar. Puedes encontrar su página aquí.

Guido es un admirador de Linux desde hace bastante tiempo. A él le gusta Linux porque te brinda libertad y oportunidades de elección. Puedes elegir y desarrollar soluciones de acuerdo con tus necesidades.

Abstract:

¿Spam entre tu correo? El Spam en el correo electrónico (Spam E−mail) está aumentando a pasos agigantados y es un verdadero problema para casi todos. En este artículo te explicaremos qué hacer para luchar contra esta plaga.

¿Qué es el spam−mail?

Spam−mail se presenta con muchos nombres. Algunos lo llaman UCE (Unsolicited commercial email) otros lo llaman simplemente Unwanted E−mail (correo no deseado) pero ninguno de estos nombres te dice qué es realmente. Si (aún) no has recibido spam entonces puedes ver this collection of spam−mail (spam_samples.html). Es una selección tomada al azar de spam−mail recibido sólo en unos días. Lee los mensajes y pronto descubrirás que no tiene nada que ver con temas comerciales o de negocios. Todo esto es delictivo. Ningún hombre o mujer de negocios serio molestaría a millones de personas para hallar a unos pocos "idiotas" que les compren sus trucos.

Existe una confusión muy común entre la gente que no ha usado demasiado Internet. Piensan que estos avisos publicitarios se pueden comparar a la información que de tanto en tanto reciben de su supermercado local. A menudo los productos que se venden vía spam−mails son ilegales o no son productos en absoluto. Son trucos para obtener tu dinero.

¿Cúanto?

Quienes envían Spam por correo electrónico (Spammers) obtienen tu dirección de correo de páginas web, grupos de discusión (news groups) o registros de dominios (en caso de que tengas uno propio). Hay personas que utilizan robots para conseguir direcciones, las queman en CDs y las venden a muy bajo precio a otros Spammers. Si escribes tu dirección de correo en el texto de tu página hoy esos programas la pueden extraer, y entonces te encontrarás con un problema mayor dentro de unos meses que no podrás detener. ¡El problema aumentará día a día!

En 1998 el porcentaje de spam mail enviado a LinuxFocus fue menor al 10%. A partir del mes de noviembre de 2002 las estadísticas son las siguientes:

Nuestro servidor recibe aproximadamente 4075 mensajes por semana. ¡3273 son spam−mails!
=> El 80% de todo el correo es Spam.

Esto significa el 80% de la capacidad del servidor y el 80% de un espacio utilizado en algo que nadie quiere.

De estos 3273 spam mails (mensajes con spam) alrededor del 40% se origina en Norte América (Principalmente en Canada, US, Mexico) y alrededor de un 30% en Asia (principalmente en Korea, China, Taiwan).

Qué hacer con el Spam

Si observas spam−mails verás que casi todos te ofrecen la posibilidad de ser removido de la lista. ¡No lo hagas! Estás tratando con delincuentes. Ningún spammer consigue nada si tiene una lista de remoción apropiada. ¿Por qué agregan esta posibilidad? La respuesta es simple. Produce una mejor impresión en el lector y es una excelente herramienta estadística. Los spammers pueden chequear de inmediato que sus mensajes de correo llegan. En otras palabras **¡tú les confirmas la recepción del mensaje!**

También hay un problema técnico simple con la lista de remoción. LinuxFocus no es un sitio muy grande pero nosotros necesitaríamos a una persona que trabajara tiempo completo para eliminar de la suscripción a 3273 Spam mails por semana y luego esta persona debería eliminar mensajes por minuto. Cada spammer utiliza un método diferente, por lo cual, sería una idea tonta y no funcionaría. Remover listas no tiene sentido y a los únicos que ayuda es a los spammers.

Lo único que debe hacerse es: borrarlo.

Software para manejar el spam

Hay muchas opciones para filtrar el spam y esto es bueno porque a los spammers les resulta más difícil evadir esto. Sin embargo, se trata de una verdadera lucha. Las herramientas para filtrar spam son cada vez más sofisticadas pero los spammers mejoran sus métodos también.

Existen 2 tipos de filtros:

1. Chequear directamente en el MTA (Message Transfer Agent=Mail server). Aquí generalmente puedes rechazar el correo. Es decir: ni siquiera lo guardas. Envías un código de error tan pronto como reconoces que se trata de spam mientras se realiza la recepción del correo. Herramientas típicas de este tipo son IP based blocklists y mail header checks.

2. Filtrar después de recibir el correo. En este caso el correo se envía con éxito y se lo filtra más tarde.

Ahora discutiremos las distintas posibilidades en detalle, todas ellas presentan ventajas y desventajas. La mejor solución para liberarse de todo spam es utilizar distintos tipos de herramientas.

Rechazar correo directamente en el MTA

Si rechazas el correo directamente en el servidor de correo durante la recepción del mismo entonces el spammer puede recibir un código de error y sabe que esta dirección no funciona. Si es un "hacedor de CD", podría quitar la dirección. Esto salva espacio ya que no tienes que recibir el mensaje completo. Puedes enviar el error tan pronto como descubres que se trata de spam.

Para hacer esto necesitas un buen MTA, que sea flexible. Desafortunadamente los dos servidores más comunes, Sendmail y el de Bill Gates no son buenos en absoluto para esto. Dos muy buenas alternativas son Postfix y Exim. Si no puedes cambiar tu servidor puedes colocar un smtp proxy como messagewall frente al servidor (smtp = Simple Mail Transfer Protocol, el protocolo de Internet mail).

Ahora discutiremos algunas técnicas comunes de filtrado y cómo funcionan. No describiremos cómo configurarlas en cada MTA. Nuestro artículo sería interminable. En cambio, sugerimos leer la documentación que trae el MTA que hayas instalado. Postfix y Exim tienen buena documentación

- Realtime Block lists:
  Son listas basadas en DNS. Chequeas la dirección IP del servidor de correo que quiere enviarte correo a tu servidor contra una lista de spammers conocidos. Encuentras listas comunes en www.spamhaus.org o en ordb.org. También hay una herramienta llamada blq (ver referencias) para investigar estas listas manualmente y verificar si una dirección IP se encuentra registrada en las listas. Sin embargo no deberías alegrarte demasiado con ello y elegir bien las listas ya que hay algunas que bloquean rangos completos de IP ranges simplemente porque un spammer utilizó una conexión dialup desde este ISP en algún momento. Personalmente utilizaríamos ordb.org para eliminar correo proveniente de servidores mal administrados.
  La experiencia demuestra que estas listas bloquean entre 1%−3% del correo spam.

- 8 caracteres en la línea de asunto (o subject line):
  Alrededor del 30% del spam se origina en China, Taiwan o en otros países asiáticos estos días. Si estás seguro de que no puedes leer chino entonces puedes rechazar mensajes de correo que tengan 8 caracteres (no ASCII) en el asunto. Algunos MTAs tienen una opción de configuración separada para esto pero también puedes utilizar una expresión regular que combine en el encabezado:

  /^Asunto:.*[^ −~][^ −~][^ −~][^ −~]/

  Esto rechazará mensajes que contengan más de 4 caracteres consecutivos en la línea de asunto que no se encuentran en el rango de espacio ASCII para colocar acentos. Si no te encuentras familiarizado con expresiones regulares entonces apréndelas, las necesitarás (Ver LinuxFocus artículo 53). Both exim y postfix pueden compilarse con apoyo de expresión regular perl (Lenguaje Práctico de Extracción e Informes) (ver www.pcre.org). Perl tiene las expresiones regulares más efectivas.
  Este método es bastante bueno y elimina entre 20−30% del correo spam o spam−mail.

- Listas con direcciones "From" ("De") de spammers conocidos:
  Olvídalo. Esto funcionaba en 1997. Los spammers de hoy usan direcciones falsas o direcciones de gente inocente.

- Rechazar no emisores FQDN (Fully Qualified Domain Name) y emisores de dominio desconocido:
  Algunos spammers usan direcciones que no existen en el "De" ("From"). No es posible chequear la dirección completa pero puedes ver parte del nombre de host (hostname)/dominio e investigar un servidor DNS.
  Esto rechaza aproximadamente entre 10−15% del spam y tú no quieres estos mensajes porque no podrías responderlos aún si no fueran spam.

- La dirección IP que no tiene registro PTR en el DNS:
  Esto chequea que la dirección IP desde la cual tú recibes el mensaje pueda ser convertida en un nombre de dominio. Esta es una opción muy efectiva y rechaza bastante correo. ¡Nosotros no la recomendaríamos! No chequea si el administrador de sistema del servidor de correo es bueno sino si tiene un buen proveedor de red vertical. ISPs compra direcciones IP de sus servidores de red vertical y le compran a servidores de red vertical mayores. Todo esto involucra a proveedores de red vertical y los ISPs tienen que configurar sus DNS correctamente para que toda la cadena funcione. Si alguien en el medio comete un error o no quiere configurarlo entonces no funciona. No dice nada con respecto al servidor de correo individual que se encuentra al final de la cadena.

- Requerir comando HELO:
  Cuando 2 MTAs (mail servers o servidores de correo) se comunican entre sí (vía smtp) primero dicen quiénes son (por ejemplo, mail.linuxfocus.org). Algunos programas de software de spam no lo hacen. Esto rechaza entre 1−5% del spam.

- Requerir comando HELO y rechazar servidores desconocidos:
  Tomas el nombre que obtienes en el comando HELO y luego vas a DNS y chequeas si es un servidor correctamente registrado. Esto es muy bueno ya que un spammer que usa una conexión telefónica temporaria generalmente no configurará un registro DNS válido para ello.
  Esto bloquea aproximadamente 70−80% del spam pero también rechaza correo legítimo que proviene de sitios con múltiples servidores de correo en donde un administrador de sistema olvidó poner los nombres de todos los servidores en el DNS.

Algunos MTAs presentan aún más opciones pero las que vimos anteriormente suelen estar disponibles en un buen MTA. La ventaja de todos estos chequeos es que no son intensivos en la CPU. Generalmente no necesitarás actualizar el hardware de tu servidor de correo si realizas estos chequeos.

Filtrar correo ya recibido

Las siguientes técnicas se aplican generalmente a la totalidad del correo y el servidor de correo que envía los mensajes no advierte que el correo podría no ser entregado. Esto significa además que el emisor legítimo no recibirá un informe de rechazo. El mensaje simplemente desaparecerá.
Habiendo explicado esto, debemos aclarar que no es completamente correcto ya que realmente depende de las posibilidades de filtro que ofrezca el servidor de correo. Exim es muy flexible y te permitiría incluir filtros personalizados en los mensajes.

- SpamAssassin (http://spamassassin.org/):
  Es un filtro de spam escrito en perl. Utiliza reglas escritas cuidadosamente y asigna ciertos puntos a frases de spam típicas tales como "strong buy" (fuerte compra), "you receive this mail because" (recibes este mensaje porque), "Viagra", "limited time offer" (oferta de tiempo limitado)... Si los puntos superan un determinado nivel el correo es declarado spam. El problema con este filtro es que es muy pesado, en términos de memoria y energía de la cpu. Probablemente necesitarás actualizar el hardware de tu servidor de correo, especialmente si el servidor ya tiene entre 2 y 3 años de

antigüedad. No recomendaríamos su uso directamente sobre el servidor de correo. Spamassassin viene con un programa spamd (spamd=spam daemon + spamc=client to connect to the daemon) que reducirá el tiempo de inicio de spamassassin así como el consumo de cpu pero es aún una aplicación que demanda muchos recursos.

Para filtrar el correo necesitas crear un .procmailrc file (y .forward) similar a éste:

```
# The condition line ensures that only messages smaller
than 50 kB
# (50 * 1024 = 56000 bytes) are processed by SpamAssassin.
Most spam
# isn't bigger than a few k and working with big messages
can bring
# SpamAssassin to its knees. If you want to run
SpamAssassin without
# the spamc/spamd programs then replace spamc by
spamassassin.
:0fW:
* < 56000
| /usr/bin/spamc
# All mail tagged as spam (e.g. with a score higher than
the set threshold)
# is moved to the file "spam-mail" (replace with /dev/null
to discard all
# spam mail).
:0:
* ^X-Spam-Status: Yes
spam-mail
```

La instalación es fácil y spamassassin filtrará más del 90% del spam.

- procmail (http://www.procmail.org):
  Procmail no es un filtro de spam por sí mismo pero puedes usarlo para escribir uno tú. procmail es además muy liviano en tanto limites el número de reglas de modo razonable (por ejemplo, menos de 10). Para utilizarlo creas un archivo del tipo .forward en tu directorio de inicio y agregas allí la siguiente línea:

  ```
  "| exec /usr/bin/procmail"
  ```

  Algunas personas recomiendan usar
  "|IFS=' ' && exec /usr/bin/procmail"
  pero esto genera nuevos problemas con la creación de un proceso extra que ya no se ejecuta bajo el control del servidor de correo. Servidores de correo seguros como postfix o exim no presentarán problemas con el archivo .forward como se mostró anteriormente.

  Procmail es especialmente útil en un entorno en donde normalmente te comunicas dentro de un grupo cerrado. Por ejemplo, para personas en una empresa en donde la mayor parte del correo proviene de colegas y algunos amigos conocidos. Aquí hay un ejemplo para "mycompany.com":

  ```
  # .procmailrc file.
  # search on header for friends:
  :0 H:
  * ^From.*(joe|paul|dina)
  /var/spool/mail/guido

  # search on header for mails which are not coming from
  # inside mycompany.com and save them to maybespam
  ```

```
:0 H:
* !^From.*(@[^\@]*mycompany\.com)
/home/guido/maybespam

# explicit default rule
:0:
/var/spool/mail/guido
```

Esto facilita la eliminación de spam y permite que ya no lo encuentres entre tu correo normal.

Procmail es muy flexible y se lo puede utilizar también para otras tareas. Aquí hay un ejemplo completamente diferente:
Procmail trae un programa "reply to sender" ("responder al emisor") llamado formail. Este programa puede utilizarse, por ejemplo, para reenviar un mensaje a la gente. Una verdadera plaga la constituyen aquellos mensajes que contienen documentos de word. Si eres un desarrollador de Linux que utilizas el correo electrónico para intercambiar información sobre tus proyectos o Linux en general, es seguro que no tendrás interés en la gente que escribe texto en un documento de word y lo anexa a mensajes. Los virus pueden propagarse rápidamente de ese modo. En general no infectan a Linux pero no es buena idea utilizar MS−word para enviar texto a otras personas ya que requiere que el receptor cuente con la misma versión de MS−word para poder leerlo. Existen formatos abiertos tales como RTF o HTML que no propagan virus, son de plataforma cruzada y no presentan el problema de contar con determinada versión.

```
# Promail script to
# reject word documents. Reject the mail, but do not reply
to
# error messages "From MAILER-DAEMON"
# If you use ":0 Bc" instead of ":0 B" then you will still
get the mail
:0 H
* !^From.*DAEMON
{
  # The mime messages with word documents look like this in
the body
  # of the message:
  #------=_NextPart_000_000C_01C291BE.83569AE0
  #Content-Type: application/msword;
  #        name="some file.doc"
  #Content-Transfer-Encoding: base64
  #Content-Disposition: attachment;
  #        filename="real file.doc"
  :0 B
  * ^Content-Type:.*msword
  | (formail -r ; cat /home/guido/reject-text-msword ) |
$SENDMAIL -t
}

# explicit default rule
:0:
/var/spool/mail/guido
```

El archivo de texto /home/guido/reject−text−msword debería contener un texto que explique que los documentos del tipo msword pueden propagar virus y solicite al emisor que envíe el documento, por ejemplo, en formato RTF.

Para saber cómo utilizar procmail y qué significan todas estas extrañas letras en el archivo de configuración encontrarás muy buenas explicaciones en la página principal de "procmailrc".

- bogofilter (http://www.tuxedo.org/~esr/bogofilter):
  Bogofilter es un filtro de spam Bayesian. Está completamente escrito en C y es muy veloz (comparado con SpamAssassin). Un filtro Bayesian es un filtro estadístico al que primero tendrás que entrenar para que identifique qué es spam y qué no lo es. Necesitas alrededor de 100 mensajes de entrenamiento (entre spam y no spam) hasta que el filtro pueda funcionar correctamente sobre tus nuevos mensajes.

  Bogofilter es rápido pero no funciona en un día como SpamAssassin. Después de un tiempo llega a ser tan eficiente como SpamAssassin y filtra más del 90% del spam.

- razor (http://razor.sf.net/):
  Es un sistema de detección de spam distribuido y colaborativo. Las sumas de verificación de mensajes de spam conocidos se almacenan en una base de datos. Si recibes correo nuevo, computas la suma de verificación y la comparas con sumas de verificación en la base de datos central. Si la suma de verificación coincide puedes rechazar el mensaje como spam. razor funciona porque en Internet se han propagado cuentas especiales de correo con el único propósito de ingresar en las listas de direcciones de todos los spammers. Estas cuentas sólo captan spam y no correo normal. La gente puede, por supuesto también, enviar mensajes a razor para transformarlo en spam. Existe una buena posibilidad que los mensajes ya sean conocidos como spam antes de que lleguen a tu casilla de correo. El sistema filtra alrededor del 80% del spam. razor presenta una característica que no tiene ningún otro proceso o técnica de filtrado: razor prácticamente no detecta falsos positivos. En otras palabras, el número de mensajes que no constituyen spam pero que aún así fueron declarados como tal se reduce significativamente con razor.

Hay muchas otras posibilidades de luchar contra el spam. Nosotros consideramos que lo que hemos expuesto anteriormente abarca las más importantes.

La mejor solución es realizar chequeos en el agente de transferencia de mensajes (o MTA) como primera etapa y luego destruir el spam remanente en una segunda etapa con un filtro post−proceso.

Mensajes del tipo HTML

Una forma particularmente peligrosa de correo la constituyen los mensajes de spam en formato HTML.

La mayoría de los spammers utilizan la "posibilidad de anular la suscripción" para verificar cúantos de sus mensajes llegan. Los mensajes en formato HTML ofrecen una forma de obtener respuesta (o feedback) mucho mejor: Imágenes. Puedes comparar este sistema con los contadores de visitas que se encuentran en algunas páginas de la red. El spammer puede ver exactamente cúando y cúantos de sus mensajes son leídos. Si estudias el spam cuidadosamente verás que en algunos casos la URL para imágenes incluidas contiene un número de secuencia: El spammer puede ver quién mira el correo y en qué momento. Un agujero de seguridad increíble.

Los programas de lectura de correo modernos no mostrarán imágenes que se descargan de algún lugar a partir de una URL. Sin embargo apenas encontrarás un lector de HTML moderno y seguro. Kmail y la última versión de mozilla mail ofrecen la posibilidad de desactivar las imágenes de recursos externos. La mayoría de los otros programas generarán lindas estadísticas para el spammer.

¿La solución? No utilices un programa que posibilite el correo html o bien descarga el correo primero, desconéctate de Internet y luego lee el correo.

¿De dónde viene el spam?

¡Nunca confíes en la dirección del emisor que aparece en la línea "De" (o "From) de los mensajes de spam! Son usuarios inexistentes o bien personas inocentes. Es muy raro que se trate de la dirección del spammer. Si deseas saber de dónde viene el correo entonces deberás ver el encabezado completo:

```
...

Received: from msn.com (dsl-200-67-219-28.prodigy.net.mx
[200.67.219.28])
        by mailserver.of.your.isp (8.12.1) with SMTP id
gB2BYuYs006793;
        Mon, 2 Dec 2002 12:35:06 +0100 (MET)
Received: from unknown (HELO rly-xl05.dohuya.com)
(120.210.149.87)
        by symail.kustanai.co.kr with QMQP; Mon, 02 Dec
2002 04:34:43
```

He aquí un host desconocido con dirección IP 120.210.149.87 quien reclama ser rly−xl05.dohuya.com que envía el correo a symail.kustanai.co.kr. symail.kustanai.co.kr envía este mensaje más tarde.
El spammer se oculta en algún lugar detrás de 120.210.149.87 que es probablemente sólo una dirección telefónica dinámica IP.

En otras palabras, la policía podría encontrar a esta persona si se dirigieran al responsable de kustanai.co.kr y solicitaran los registros del servidor y una impresión de las conexiones en la compañía de teléfono local. Tú tienes pocas posibilidades de descubrir quién es.

También podría suceder que la primera parte fuera falsa y que el spammer se escondiera detrás de dsl−200−67−219−28.prodigy.net.mx. Esto es muy probable ya que no existe una buena razón para que symail.kustanai.co.kr le enviara correo a msn.com via la conexión telefónica dsl (dsl−200−67−219−28.prodigy.net.mx). El servidor de correo.de.tu.isp (nombre simbólico) es el servidor de tu Proveedor de Internet y solamente la parte de esta línea de "Recibido" (o "Received")es confiable.

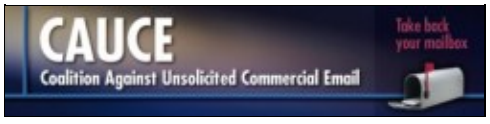
Es posible hallar al spammer pero necesitas inteligencia international y fuerzas de seguridad para llegar a prodigy.net.mx.

Conclusión

Si el spam continúa creciendo a la velocidad que lo está haciendo, Internet pronto transportará mucho más Spam que verdadero correo. La circulación del Spam la paga el receptor. Se necesita mayor ancho de banda y a menudo los sistemas de correo deben ser actualizados para poder manejar el Spam.
En muchos países las leyes no ayudan mucho a la gente para protegerla contra los delincuentes spammers. De hecho algunos países cuentan con leyes que restringen únicamente a la gente honesta (administración de derechos digitales, etc. ...) y ayudan a los delincuentes (por ejemplo, a conseguir lindas estadísticas acerca del correo spam).

¡Únete a la coalición contra el UCE (o Correo Comercial No Solicitado)!

http://www.euro.cauce.org/en/



http://www.cauce.org/

Los proveedores de servicios de Internet deberían chequear sus sistemas de correo. No debería permitirse el acceso irrestricto a los servidores de correo y debería limitarse el número de mensajes que un usuario puede enviar por minuto.

Referencias

- http://spamassassin.org/: página de inicio de spamassassin
- http://www.procmail.org/: página de inicio de procmail
- http://www.spambouncer.org/: spambouncer: a procmail based spam filter
- http://www.postfix.org/: página de inicio del MTA de postfix
- http://www.exim.org/: página de inicio del MTA exim
- http://messagewall.org/: página de incio de messagewall smtp proxy
- http://www.unicom.com/sw/blq/: the blq perl script to query DNS based block lists
- http://www.ordb.org/: DNS based open relay block list
- http://www.spamhaus.org: DNS based block list
- http://www.samspade.org/: ¿De dónde viene el spam?
- http://www.dnsstuff.com/: various blocklists and DNS based tools
- http://www.geektools.com/cgi−bin/proxy.cgi: geektools Whois proxy
- http://www.tuxedo.org/~esr/bogofilter/: filtro de correo bogofilter
- http://razor.sf.net/: razor
- http://pyzor.sourceforge.net: razor implementado en python
- http://lwn.net/Articles/9460/: el artículo de las noticias semanales de Linux en el que se comparan bogofilter y spamassassin.

2005−01−10, generated by lfparser_pdf version 2.51