Universidad Central «Marta Abreu» de Las Villas

Facultad de Ingeniería Eléctrica

Departamento de Automática y Sistemas Computacionales

TRABAJO DE DIPLOMA

Proposición de estrategias para contrarrestar el efecto del SPAM en la Intranet de la UCLV

Autor: Raydel Apesteguia Ojeda

Tutor: Manuel Oliver Domínguez

Santa Clara

2007

«Año 49 de la Revolución»

Universidad Central «Marta Abreu» de Las Villas

Facultad de Ingeniería Eléctrica

Departamento de Automática y Sistemas Computacionales

TRABAJO DE DIPLOMA

Proposición de estrategias para contrarrestar el efecto del SPAM en la Intranet de la UCLV

Autor: Raydel Apesteguia Ojeda

raykuba@gmail.com

Tutor: MSc. Manuel Oliver Domínguez

Profesor Departamento de Automática. Facultad de Ingeniería Eléctrica. Universidad Central «Marta Abreu» de Las Villas. moliver@uclv.edu.cu

Consultante: MSc. Iván Iglesias Navarro.

Santa Clara

2007

«Año 49 de la Revolución»

Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central «Marta Abreu» de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería en Automática, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

_____

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

_____ _____

Firma del Autor

Firma del Jefe de Departamento donde se defiende el trabajo

_____

Firma del Responsable de Información Científico-Técnica

«…el hombre ansía siempre una felicidad situada más allá de la porción que le es otorgada. Pero la grandeza del hombre está precisamente en querer mejorar lo que es.

En imponerse Tareas.»

Alejo Carpentier

Dedico esta tesis a los que no se dan por vencidos,

a mi familia, en especial a mis padres y hermana, por su confianza.

A todos lo que han hecho posible este sueño.

A mi mamá y a mi papá que tuvieron paciencia para guiarme siempre.

A mi único abuelo, que tanto me quiere.

A mis profesores por enseñarme siempre lo correcto, entre ellos Iván y Robby que al fin van a descansar, los llevo conmigo.

A mi tutor Manuel «el volao», mejor no lo quiero.

A Melquíades y Lester por su ayuda, gracias.

A mis amigos de siempre, ya lo logré.

A mi tía Dulce y a Magdie que también están aquí, como los quiero.

A Yai.

Y a Rosi, por extrañarme.

TAREA TÉCNICA

1. La revisión bibliográfica relacionada con el correo electrónico y el SPAM.
2. La identificación las causas, efectos y características del SPAM.
3. El análisis crítico de las técnicas anti SPAM.
4. La confección de un informe que contenga los principales resultados.

Firma del Autor

Firma del Tutor

RESUMEN

El envío de SPAM ha pasado de ser un problema «molesto» a un grave problema para el correcto funcionamiento del servicio de correo en Internet. Muchas organizaciones como la Universidad Central «Marta Abreu» de Las Villas y usuarios domésticos ven afectado su rendimiento, pues su ancho de banda de conexión se ve limitado, además de emplear tiempo diario para eliminar estos mensajes no deseados de sus buzones. En la actualidad, se han desarrollado diferentes técnicas anti SPAM pero el problema aun no se ha erradicado por completo, pues lo *spammers* están en constante desarrollo, convirtiendo al SPAM en un arma que evoluciona a diario. Para intentar evitar y atenuar este problema en las redes conectadas a Internet y en especial en la red de la UCLV es necesario conocer el origen del mismo, su evolución y algunas de las técnicas o combinación de técnicas utilizadas para evitarlo y combatirlo. Las soluciones propuestas se basan en software libre y accesible a todos. El EXIM y el *spamassassin* son el eje central del mecanismo de control propuesto.

TABLA DE CONTENIDOS

INTRODUCCIÓN

En la actualidad, tener un buzón de correo implica automáticamente recibir publicidad. Da igual que el buzón sea físico o electrónico. Sin embargo, el bajo costo del envío de los mensajes de correo electrónico y el enmascaramiento de su remitente ha dado lugar a una verdadera invasión de correos electrónicos no solicitados, también llamados SPAM.

Este fenómeno es considerado hoy un grave problema de las redes conectadas a Internet. Proporcionando pérdidas desde todas las aristas para los acreedores del servicio, pues además del consumo del ancho de banda de conexión implica tiempo perdido, aspectos que influyen poderosamente en la productividad de una empresa o usuario. Por tanto, es indispensable para cualquier usuario, institución o empresa trazar e implementar estrategias y herramientas que permitan, sino la eliminación total, al menos la eliminación parcial del mismo.

A nivel mundial combatir el SPAM es una necesidad, pues la nueva era incluye el SPAM gráfico, el político y los asociados a fraudes acentuando aun más sus perjudiciales características. Subrayando que este fenómeno está en constante desarrollo por la evolución que experimenta a diario haciéndose más difícil lograr identificarlo por la inclusión de nuevas técnicas para evadir los filtros anti SPAM.

La red de la Universidad Central «Marta Abreu» de Las Villas, producto a su conexión a Internet no ha quedado exenta del problema, atentando así contra su óptimo rendimiento y evidenciándose todos los efectos que trae consigo el SPAM.

Luego de realizar un análisis de los recursos que este fenómeno consume surgió la necesidad de combatir el SPAM esta institución, pues en la UCLV no se existía

ninguna herramienta para atenuar el problema convirtiéndose esta problemática en unos de los pilares de este trabajo de diploma.

La atenuación de la incidencia del SPAM son las bases que sustentan a la presente tesis de pregrado. Trazándose como objetivo general el desarrollo de estrategias para contrarrestar el efecto del SPAM en redes conectadas a Internet. En correspondencia con este objetivo se presentan los objetivos específicos siguientes:

- Revisar bibliografía relacionada con el correo electrónico y el SPAM.
- Identificar las causas, efectos y características del SPAM.
- Analizar críticamente las diferentes técnicas anti SPAM.

Las fuentes bibliográficas consultadas durante la realización del proyecto abarcan preferentemente aquellas soluciones anti SPAM desarrolladas en los últimos años para servidores Exchange y UNIX[1], los cuales están presente en la red UCLV. Se abarca lo relativo a la excelente integración del Exim4 con *spamassassin*. Se efectúa en ese sentido un riguroso estudio documental de las fuentes bibliográficas, fusionándose el método analítico-sintético y el empírico mediante la práctica de la observación.

Organización del informe

El objetivo general y los objetivos específicos se desarrollan mediante una estructura lógica que basada en:

- Introducción.
- Capítulo I Correo y SPAM: pasado y presente.
- Capítulo II Herramientas y estrategias para el control del SPAM.
- Capítulo III Situación red UCLV.

[1] En este trabajo los sistemas operativos Linux quedan incluidos siempre que se mencione al UNIX.

- Conclusiones.
- Recomendaciones

En el capítulo I «Correo y SPAM: pasado y presente», luego de realizar una búsqueda bibliográfica sobre el correo y el SPAM de ver las causas que lo originan, las consecuencias para Internet y algunas formas de evitarlos se crea un texto que permitirán obtener la cultura necesaria para enfrentar el problema planteado.

El capítulo II «Herramientas y estrategias para el control del SPAM», plantea las distintas estrategias tanto globales como locales que minimizan el SPAM, analiza las diferentes técnicas para el filtrado y realiza un análisis de los productos anti SPAM así como sus proveedores, haciendo énfasis en la solución óptima para el correcto funcionamiento de un servidor de correo y la atenuación del SPAM.

En el capítulo III «Situación red UCLV», primeramente se hace una análisis de la red de la universidad, luego la proposición e implementación de la solución anti SPAM lograda y por último el análisis económico que indica la rentabilidad del proyecto ejecutado.

Finalmente, la totalidad de este trabajo pretende que se comprenda mejor el problema del SPAM y lo difícil que es combatirlo y eliminarlo. Se muestra que se pueden lograr soluciones muy prometedoras siempre y cuando se mantenga la supervisión y el control de este problema.

CAPÍTULO 1. CORREO Y SPAM: PASADO Y PRESENTE

Una de las aplicaciones más importantes de este fenómeno mundial que es Internet es el correo electrónico. Quizás sea la primera que llegue a la mente de la mayoría de las personas que usan la red de redes. Cada día más personas saben lo que es una dirección de correo electrónica, que es un "*e-mail*" y como se envía y se recibe.

En este capítulo abordaremos este tema desde un punto de vista técnico pero sin olvidar los principales hechos y situaciones que nos llevan a nuestro problema principal: el SPAM.

1.1 Correo electrónico

El correo electrónico, también llamado e-mail (*electronic mail*), es una forma de enviar mensajes que contengan no solo texto entre personas usando un computador u otro dispositivo compatible según (Chávez, 2000). Estos mensajes o «cartas electrónicas» se escriben en una computadora local y se envían a través de las redes de computadoras a sus destinatarios, quienes junto al remitente deben disponer de una dirección de correo electrónica válida en un servidor de correo.

En realidad el correo electrónico es muy parecido al correo tradicional, aunque tiene varias diferencias, en la tabla 1.1 se puede observar esto:

Tabla 1.1. Comparación entre el correo tradicional y el correo electrónico

Servicio	Costo	Tiempo	Copias	Información
Correo tradicional	se paga cada mensaje (sellos, papel, sobres)	algunos días	se realiza una a una cada copia	solo textos
Correo electrónico	servicio privado y gratuito	minutos a cualquier parte del mundo	con solo indicar las direcciones	textos, fotos, ejecutables etc.

Como se aprecia, el correo electrónico es mucho más versátil que el correo tradicional pero su mayor ventaja es que logra enviar gran volumen de información en muy poco tiempo y con mayor fiabilidad.

El servicio de correo fue de los primeros que ofreció la red ARPANET ver (Menéndez, 2004). Al principio el sólo enviaba mensajes de texto con mayor o menor rapidez entre pocos usuarios. En la actualidad, es posible enviar todo tipo de datos binarios, permitiendo incluir como parte del mensaje imágenes, sonidos, ficheros binarios, programas y ejecutables.

Para poder usar el servicio de correo electrónico hacen falta tres elementos:

- Un cliente que origine un correo usando una herramienta adecuada que cumpla los estándares existentes.
- Uno o varios servidores que acepten, encaminen y entreguen el correo.
- Un destinatario que tome el correo desde su servidor y lo muestre usando un programa similar y completamente compatible con el usado para enviarlo.

Un servidor de correo técnicamente no es más que un programa que está en escucha en un ordenador por el puerto 25 TCP y que cumple determinadas funciones, respondiendo a ciertos datos de entrada que recibe, ver (Osmosis,

2007). En sí, gestiona la entrada y salida de correos desde y hacia Internet o sin salir de la propia Intranet.

Estos servidores no necesariamente deben pertenecer a la red de Internet, pues se pueden enviar correos electrónicos a usuarios de otras redes a través de computadoras llamadas *mail gateways* o *relay servers* que permiten la comunicación.

Una característica importante del e-mail, es que si un usuario tiene un acceso limitado a Internet, con su cuenta de correo puede sacar mucho provecho a servicios como el FTP, listas de correos, Web entre otros.

Es muy importante saber lo que es una dirección de correo electrónico cuando hablamos de este tema, consultar (Chávez, 2000).

Una dirección de correo electrónico es la forma que tenemos de especificar al programa de correo electrónico, el lugar o persona a la que queremos enviar el mensaje en concreto. La dirección de correo electrónico tiene la siguiente forma:

usuario@pcorigen.dominio

Primero se pone el usuario, indica el buzón de correo electrónico correspondiente a la persona a la que va destinado el mensaje. Después se coloca un símbolo que se denomina arroba y a continuación el nombre del dominio donde tiene cuenta el usuario.

Hay varios tipos de dominios en Internet. Normalmente suelen tener como máximo 3 letras que los identifican, como los asociados con cada país, ver (Graphics, 2007)

1.1.1 Protocolos que intervienen en una aplicación de correo electrónico

El Protocolo Simple de Transferencia de Correo (*Simple Mail Transfer Protocol, SMTP)* es un protocolo estándar básico de Internet del nivel de aplicación utilizado

para la transmisión de correo electrónico a través de una conexión TCP/IP, opera sobre el puerto 25 y es basado en texto. Constituyó el principal protocolo utilizado para la transmisión de correo electrónico a través de Internet pero como no permitía anexar archivos, soportaba solo ASCII de 7 bits y sus líneas no excedían los 1000 bytes, fue necesaria su actualización. En la actualidad se utiliza el Protocolo Simple de Transferencia de Correo Extendida (*Extended Simple Mail Transfer Protocol, ESMTP)*. Este nuevo protocolo que extiende los servicios de *SMTP*, permite anexar archivos, soporta diferentes tipos de contenidos, y define los siguientes tipos de codificación:

- **8 BitTime**: Datos binarios de 8 bits.
- **Chuncking**: Mensajes en partes o trozos (*chunks*)
- **Check Point**: Soporta transacciones de correo.
- **Pipelining**: Múltiples comandos en un solo envío.
- **Size**: Muestra el máximo tamaño de los mensajes que acepta.
- **DSN**: Proporciona notificaciones de estado de entrega de mensajes.
- **ETRN**: Acepta solicitudes remotas de procesamiento en cola.
- **Enhanced Status Codes**: Muestra una lista con los códigos de errores mejorados

Para encontrar el servidor *SMTP* de un dominio dado, se utilizan los registros intercambiadores de correo (*Mail Exchanger, MX*) en la zona de autoridad correspondiente a ese mismo dominio contestado por un servidor de *DNS*. Los registros *MX*, también pueden manejar el tráfico de anfitriones que no están conectados a la red, como UCCP o FidoNet, que necesitan un *gateways* de correo.

Este protocolo está compuesto por un grupo de comandos esenciales que serán descritos a continuación pero primero se debe analizar lo que sería una sesión común en un servidor *SMTP* o de correo. El ejemplo muestra el uso de la

aplicación telnet para enviar un correo.

Cliente: $ **telnet 127.0.0.1 25**

Servidor: Trying 127.0.0.1... Connected to localhost.localdomain(127.0.0.1). Escape character is '^]'. 220 nombre.dominio ESMTP Sendmail 8.13.1/8.13.1; Sat,18 Mar 2007 16:02:27 -0600

Cliente: **HELO localhost.localdomain**

Servidor: 25 nombre.dominio Hello localhost.localdomain[127.0.0.1], pleased to meet you

Cliente: **MAIL FROM:fulano@localhost.localdomain**

Servidor: 250 2.1.0 <fulano@localhost.localdomain>... Sender ok

Cliente: **RCPT TO:root@localhost.localdomain**

Servidor: 250 2.1.5 <root@localhost.localdomain>... Recipient ok

Cliente: **DATA**

Servidor: 354 Enter mail, end with "." on a line by itself

Cliente:

Subject: Mensaje de prueba

From: fulano@localhost.localdomain
To: root@localhost.localdomain
Hola. Este es un mensaje de prueba.
Adios.
.

Servidor: 250 2.0.0 k2IM2RjA003987 Message accepted for delivery

Cliente: **QUIT**

Servidor: 221 2.0.0 nombre.dominio closing connection

Servidor: Connection closed by foreign host.

Ahora una breve explicación de esta secuencia.

HELO hostname:

Abre una sección con el servidor de correo, permitiendo al cliente identificarse a sí mismo. El *ESMTP* adiciona el comando: *EHLO* y las respuestas de dicho comando.

MAIL FROM direccion@decorreo.com

Indica el remitente del mensaje.

RCPT TO direccion@remota.com

Indica los destinatarios del mensaje.

DATA

Le indica al servidor de correo que todo lo que va detrás de él, es el cuerpo del mensaje. Es importante destacar, que el final del mensaje queda especificado con (354 End data with ¨.¨) que da por terminado el mismo una vez que encuentre un punto solo en una línea a parte.

QUIT

Cierra la sección

La descripción completa de los protocolos *SMTP* y *ESMTP* están definidas en el RFC 821 y el RFC 2821 respectivamente, según (Society, 2001)

1.1.2 Correo electrónico y la resolución de nombres

El sistema de resolución de nombres (*Domain Name System*, *DNS*) ver (Querétaro, 2006) se utiliza para distintos propósitos descritos por (Smaldone, 2006). Los más comunes son:

- Resolución de nombres: Dado el nombre completo de un *host* (por ejemplo *blog.smaldone.com.ar*), obtener su dirección IP (en este caso, *208.97.175.41*).
- Resolución inversa de direcciones: Es el mecanismo inverso al anterior. Consiste en, dada una dirección IP, obtener el nombre asociado a la misma.
- Resolución de servidores de correo: Dado un nombre de dominio (por ejemplo *gmail.com*) obtener el servidor a través del cual debe realizarse la entrega del correo electrónico (en este caso, *gmail-smtp-in.l.google.com*).

Normalmente los usuarios de correo electrónico redactan sus mensajes usando un cliente de correo y enviándolo a través de un servidor *SMTP* provisto por su proveedor de servicios de Internet (*Internet Service Provider, ISP)* o a través de un sistema de correo vía Web *(*webmail*).* En cualquier caso, una vez que el mensaje es recibido por el servidor, debe ser entregado al destinatario y aquí interviene el sistema *DNS*:

1. El servidor emisor solicita al *DNS* la entrada *MX* del dominio del receptor del mensaje. El *MX* busca el nombre del servidor o los servidores encargados de recibir los mensajes destinados a determinado dominio.
2. El *DNS* devuelve el nombre completo de un host (*Fully Qualified Host Name, FQHN*): y la dirección IP del mail Exchange. El *FQHN* esta formado por el *hostname*, seguido de un punto y su correspondiente nombre de dominio.
3. El servidor del emisor se conecta al puerto 25, mediante TCP**,** del servidor del destinatario y entrega el mensaje según el protocolo *SMTP*.
4. El proceso podrá continuar si el servidor receptor del mensaje no es el

último de la cadena, pues existen servidores que actúan como puertas de enlace o *gateways* de correo electrónico, y que se encargan de recibir los mensajes de determinados dominios para luego enviarlos a otros servidores.

1.2 ¿Qué es el SPAM?

El desarrollo alcanzado por la Internet trajo consigo un mayor volumen de intercambio de información, así como mejoras en las técnicas para este intercambio. El correo electrónico como herramienta propia también evolucionó paralelamente, aumentando el número de servidores de correos y de usuarios por servidores. El crecimiento del número de clientes con correo electrónico fue un factor que favoreció a la utilización de esta herramienta para divulgación global de información con fines tanto comercial como informativo, de aquí surgió el correo no solicitado o el llamado SPAM ver (Wikipedia, 2007c).

Originalmente SPAM se llamó al jamón con especias (*Spiced Ham*) producido por *Hormel* en 1926 como el primer producto de carne enlatada que no requería refrigeración según (Panda, 2007). Esta característica hacía que estuviera en todas partes, incluyendo en los ejércitos americanos y rusos de la segunda guerra mundial. El uso del término SPAM para relacionarlo con la recepción de información no deseada surge probablemente en Internet en los años ´80 según (Templeton, 2004a) y (MailxMail.com, 2007). Proviene de un famoso montaje del grupo cómico inglés *Monty Python*, en el que unos vikingos dentro en una tienda acompañan con una molesta canción la insistencia del propietario de que toda la comida que proporciona está hecha con SPAM. Tal vez, por esta razón, se ha utilizado el término para calificar el correo electrónico no solicitado, convirtiéndose en una de las mayores molestias para las personas en la red.

Contrario al correo basura o *junk mail* que recibimos en nuestros buzones ordinarios (físicos, en papel), el recibo de correo por la red le cuesta a un buen número de personas, tanto en la conexión como en el uso de la red misma. El

SPAM representa pérdidas directas al destinatario además del tiempo empleado para eliminarlos.

Inicialmente Internet no permitía su uso comercial, pero es el uso comercial el que sostiene su infraestructura. Por lo tanto, la definición en ese entonces se refería exclusivamente a «correo no solicitado» y tenia implícito que no era comercial. Hoy, con el uso comercial de la gran red, este correo no solicitado se puede separar en el comercial, o sea, el que quiere venderle algo; y el informativo, que le informa de un evento u ofrecimiento que no implica una pérdida económica para el receptor.

La palabra «SPAM» aplicada al e-mail significa Correo Electrónico Masivo No Solicitado (*Unsolicited Bulk Email, UBE*) según (Project, 2007)

- No solicitado significa que el Receptor no dio un permiso verificable para que se le envíe el mensaje.
- Masivo significa que el mensaje es enviado como parte de una colección mayor de mensajes, donde todos tienen el contenido sustancialmente idéntico.

Entonces, una definición técnica de SPAM sería:

Un mensaje electrónico es SPAM si:

- La identidad personal del receptor y el contexto son irrelevantes porque el mensaje es igualmente aplicable a muchos otros receptores potenciales.
- No se puede verificar que el receptor haya dado un permiso deliberado, explícito, y aún, revocable para que el mismo sea enviado.
- La transmisión y la recepción del mensaje parece a juicio del receptor dar un beneficio desproporcionado al remitente.

1.2.1 Evolución del SPAM en Internet

Existen fuentes que citan el comienzo del SPAM en el año 1994 según

(Padrón, 2007). El 12 de abril de dicho año un abogado estadounidense y su esposa, utilizaron un programa informático para ofrecer los servicios de su bufete a seis mil foros de Internet (grupos de USENET) algo que molestó a varios millones de usuarios. Como consecuencia por los ataques fueron desconectados de Internet y posteriormente, perdieron su licencia de abogados según (Wikipedia, 2007a).

En realidad, es posible que el primer SPAM generado lo fuera en 1978, cuando Gary Thuerk un comercial de la empresa DEC ver (Fernández, 2005), envió un mensaje anunciando sus nuevos sistemas informáticos. Entonces Internet era aún Arpanet según (Hauben, 2004), y el envío de un mensaje a casi seiscientas personas, también causó malestar generalizado, si bien éste tenía interés para algunos de los que lo recibieron afirma (Templeton, 2004b).

El correo no solicitado, aún sin existir estadísticas muy fiables, se estima que ha ido creciendo de forma exponencial ver (Cortés, 2002), siendo, hoy día un problema acuciante para muchas personas y organizaciones según (Chacón, 2006). Algunas personas que empezaron a enviar de forma masiva, tras ser reprendidos, se dieron cuenta de su error, otras se han acabado convirtiendo en profesionales del método, conocidos como *spammer*s, que hacen uso de todo tipo de tácticas para asegurar que son capaces de enviar correo a todos los puntos del planeta. Estos «profesionales», incluso, ofertan sus servicios a compañías privadas. Según entre el 45% y el 60% del correo recibido en Internet en el año 2003 ha sido SPAM y la situación no hace sino agravarse con el paso del tiempo como se muestra en la figura 1.1 donde se representa un gráfico con la evolución de este fenómeno desde diciembre del 2002 a mayo del 2004 según (Fernández, 2005).

El SPAM, en Internet es un problema global, ya que tanto los sistemas que generan el correo basura, como los responsables de su envío, están distribuidos por todo el mundo, no existiendo aun medidas capaces de eliminarlo por completo, aunque ya fuera un problema reconocido desde

1975 por Jon Postel, cocreador del correo electrónico, (Postel, 1975).

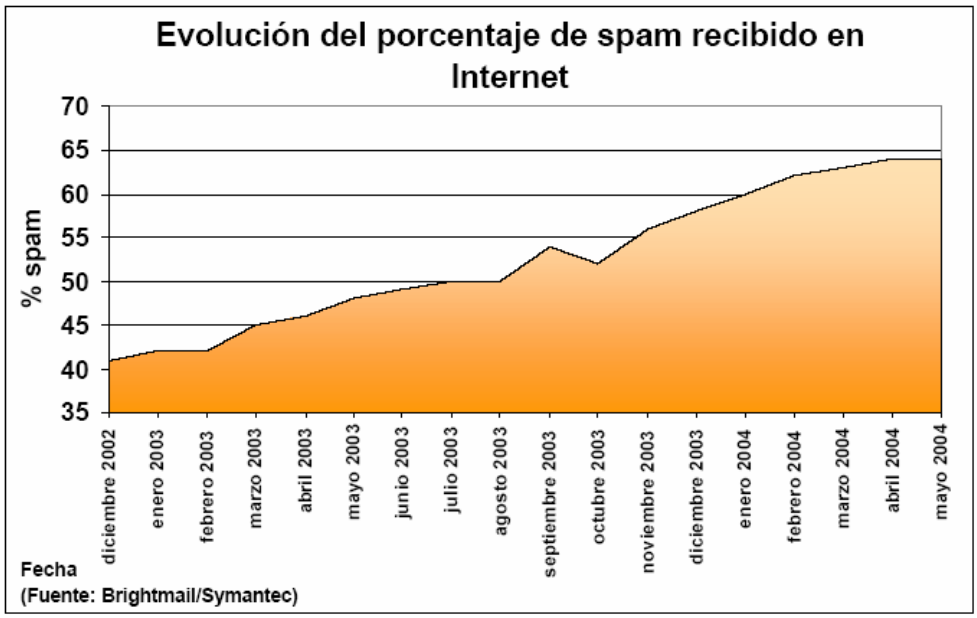


Figura 1.1. Evolución del porcentaje de SPAM recibido en Internet.

1.2.2 Tamaño del SPAM

El tamaño de los mensajes SPAM, al igual que el de los mensajes legítimos, puede variar como se muestra en la figura 1.2. Según (Kalinin yVlosava, 2007).

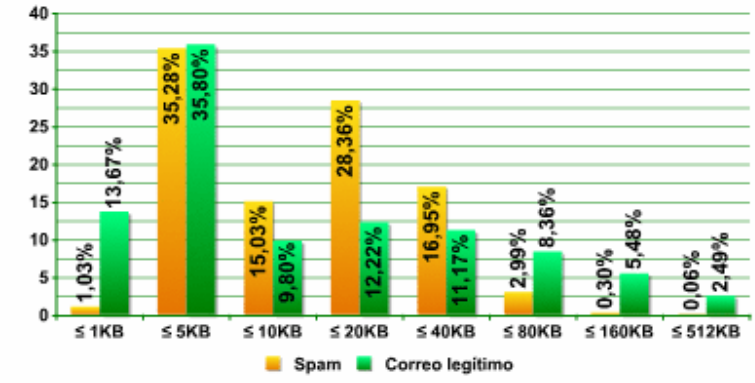


Figura 1.2 Tamaño promedio del correo en Internet.

En Internet los mensajes, tanto SPAM como legítimos, son en su mayoría (un 35%) de entre 1 a 5 KB. Además, en la distribución del tamaño en el intervalo de 10 a 20 KB de los mensajes SPAM se observa un pico elevado (28,4%) con respecto al de los mensajes legítimos y la razón está en que gran cantidad de

esos mensajes SPAM, corresponden al llamado «SPAM gráfico». En realidad el tamaño de la gran mayoría de mensajes SPAM (95,6%), está en el intervalo entre 5 y 40 KB, pues de los mensajes que no superan 1 KB (1% es SPAM) y los mensajes pesados, el tamaño es mayor a 80 KB (3,4% es SPAM).

1.2.3 Objetivos del SPAM

Si bien algunos mensajes no deseados tienen como objetivo la difusión de mensajes filosóficos, políticos o incluso religiosos enviados por personas comprometidas con una causa. Éstos suponen hoy día una pequeña parte del correo no deseado, ya que la gran mayoría del SPAM tiene fines lucrativos. La figura 1.3 da una idea de los temas más frecuentes usados por los envíos de SPAM según (Kalinin yVlosava, 2007).

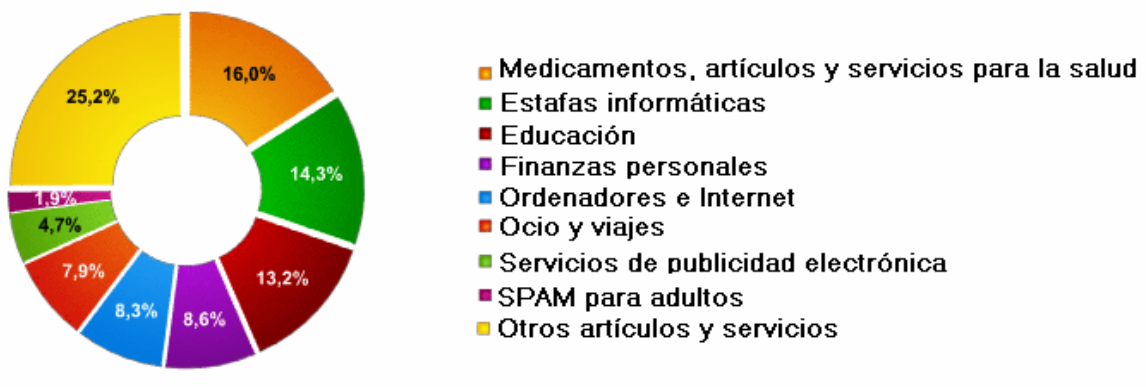


Figura 1.3. Temas más frecuentes usados por los envíos de SPAM.

Claramente, son los fines lucrativos los que mueven a los *spammers* a enviar cantidades ingentes de correo, pero, ¿realmente vale la pena? La respuesta es sí, sólo basta hacer unos cálculos muy sencillos. El costo de enviar un correo no solicitado para el *spammer* es muy bajo, casi nulo si se usan técnicas ilegales y la venta de un producto que le interesa a un porcentaje muy pequeño, digamos, un 0.1% de la población puede dejar un determinado margen, digamos que dos euros. Más aún si lo que se venden son productos comunes como por ejemplo, una aspirina como si fueran productos milagrosos a precios elevados. Si el SPAM se envía a un millón de personas los beneficios de un sólo envío indiscriminado serán de dos mil euros. Parece razonable pensar que, a más direcciones, más lucro, hasta llegar al sueño de

muchos *spammers*, disponer de todas las direcciones de los usuarios de Internet del planeta.

Ciertamente, una actividad suficientemente lucrativa. Sin embargo, estas direcciones de correo y las mismas tácticas del *spammer,* pueden ser utilizadas para llevar a cabo ataques aún más lucrativos, si estos son fraudulentos o tienen fines ilegales. Por esto, recientemente se han empezado a observar ataques dirigidos a usuarios de entidades bancarias con el afán de engañarles y obtener las contraseñas de acceso a sus cuentas para así poder transferir su dinero a otras cuentas. Estos ataques, conocidos como *phishing*, han afectado a un elevado número de entidades conocidas, incluyendo servicios en Internet (*American Online, Amazon, eBay, MSN, Paypal, Yahoo!*), bancos norteamericanos (*Bank of America, Barclays, Citibank, SunTrust, U.S. Bank*), alemanes (*Deutsche Bank, Postbank*), españoles (*Banesto, Banco Pastor, BBVA, BBK, Caja Madrid*), e ingleses (*NatWest*), entre otros.

La táctica para llevar a cabo estos ataques es similar, enviar correo no solicitado a millones de direcciones, esperar que alguno de éstos use los servicios bancarios que se están suplantando y que caigan en la trampa. El crecimiento de este tipo de ataques a lo largo del año 2005 ha sido gigantesco, en su estudio, la compañía MessageLabs indica que ha detectado 18 millones de correos asociados a ataques de *phishing* en 2005.

1.2.4 Tácticas para la recolección de direcciones de correo

Los *spammers* hacen uso de distintas técnicas para obtener las direcciones de correo a las que luego envían correo de forma indiscriminada. Algunas de estas técnicas utilizan información pública y otras pueden llegar a considerarse un delito en muchos países. Las utilizadas más frecuentemente según (Fernández, 2005) son:

- Recogida de información publicada en grupos de noticias (*USENET*), servidores Web, salas de chat y listas de correo incluso suscribiéndose a las mismas. De éstos extraen no sólo direcciones de correo de usuarios (de las cabeceras de los mensajes) sino también servidores de correo o

gateways existentes en Internet.

- La utilización de diccionarios de palabras y nombres comunes contra servidores de correo para identificar direcciones válidas en éstos.
- El ataque a sistemas informáticos, a través de instrucciones remotas (habitualmente a líneas con conexión a Internet doméstica), o bien a través del envío de virus o troyanos. El objetivo es hacerse con el control del ordenador personal de un usuario para recoger información de las direcciones de correo que éste almacena en su equipo: libreta de direcciones, correos enviados y recibidos, etc.

Se han realizado varias pruebas para evaluar la capacidad de los *spammers* para obtener direcciones de correo según (Technology, 2003). Estas pruebas muestran que de las direcciones publicadas en servidores Web conocidos, del 86% al 97% son recogidas y utilizadas en un plazo inferior a un mes por los *spammers* según (Fernández, 2005). El porcentaje para el caso de direcciones de correo utilizadas para publicar información en foros de USENET es cercano al 85%. Si bien el porcentaje se reduce en el caso de direcciones publicadas en servidores Web poco publicitados (un 50% de las direcciones publicadas en páginas personales) o en otros servicios (como las bases de datos de WHOIS) que aún no están siendo aprovechados. Algunas veces los *spammers* utilizan distintos mecanismos para comprobar si las direcciones así obtenidas son válidas. Por ejemplo, una táctica habitual consiste en enviar un correo a la dirección y comprobar si el correo se recibe o «rebota» o bien se envían correos con imágenes incrustadas (que muchos programas de correo mostrarán de forma automática) cuyos nombres incluyen marcas identificativas y están disponibles en servidores en Internet. Una vez enviado el mensaje, el *spammer* analiza los registros de acceso al servidor de donde el usuario intenta recoger la imagen. Así, puede identificar fácilmente si el correo utilizado está siendo o no leído realmente. Igualmente, muchos mensajes de correo basura incluyen indicaciones del procedimiento que debe seguirse para darse de baja de la listas de correo en la que el

usuario ha sido incluido. En realidad, cuando el usuario sigue estas indicaciones, lejos de borrar su dirección, ayuda a confirmar al *spammer* que una persona efectivamente está leyendo esa dirección de correo. Los grupos que se dedican a este oscuro negocio se intercambian estas listas de direcciones. Además, las ponen a la venta (muchas veces haciendo sus ofertas a través de más correos basura), distribuyéndolas en CD-ROMs. No se han realizado muchos análisis de estos CD-ROMs, pero Rejo Zenger, investigador holandés miembro de una fundación anti SPAM, ha publicado un análisis (Zenger, 2004) que indica que, tras eliminar direcciones sintácticamente incorrectas, direcciones genéricas y duplicadas, el CD-ROM contenía alrededor de seis millones de direcciones de correo válidas. Menos direcciones de las que prometía el vendedor (poco más de la mitad) pero aún así un número muy elevado.

En la actualidad algunas de las formas utilizadas por los *spammers* para obtener direcciones de correo pueden ser evitadas:

- Mediante una correcta configuración o supervisión de los sistemas de correo, si se evita enviar información a usuarios externos de direcciones de correo legítimas con respuestas automáticas o si se controla el envío y recepción de «cadenas de mensajes» a y desde Internet.
- Con un correcto control de los ordenadores personales (o corporativos), para evitar la infección de troyanos o virus.
- Con una correcta configuración del programa utilizado por un usuario como su cliente de correo para evitar enviar información que pueda ayudar a un *spammer* incluyendo evitar la carga automática de imágenes (u otros contenidos) de servidores remotos, no enviar respuestas automáticas a correos (incluyendo los acuses de recibo, los mensajes de vacaciones o ausencias temporales).
- También es necesario concientizar a los usuarios finales de aquellas actividades que pueden convertirles en objetivo del SPAM, incluyendo

la instalación de *software* ilegal, la introducción de información personal en servidores no confiables, o la compra de productos anunciados a través del SPAM.

1.2.5 Consecuencias del spam para internet

Las consecuencias del SPAM en Internet son nefastas. Atacan a un servicio «estrella» de Internet: el correo electrónico, haciendo que su usabilidad se vea reducida, y obliga a los usuarios de éste a lidiar con el correo no solicitado que reciben, discriminando qué correo es legítimo, cuál no lo es y borrándolo de sus buzones. Al final, cuando el volumen de SPAM es insoportable, el usuario se ve forzado a abandonar la dirección de correo y utilizar otras direcciones como sus direcciones personales.

El hecho de que los *spammers* dispongan de un elevado número de direcciones de correo válidas también hace posible que aquellos individuos que tengan la intención de difundir un ataque, como pueda ser un nuevo virus, un troyano o algún otro ataque especializado (como los ataques de *phishing*) pueda extender más fácilmente el contenido malicioso a muchas más víctimas potenciales.

La difusión inicial y la propagación de estos ataques son consecuentemente mucho mayores. Esto lleva, al final, a una pérdida de confianza de los usuarios en la propia tecnología, con consecuencias sobre los servicios electrónicos, incluido el comercio, y va directamente en contra de los esfuerzos que realizan los organismos y empresas para desplazar servicios tradicionales (desde la compra de entradas de cine a la oferta de servicios de la administración pública) a Internet.

1.2.6 La batalla legal y global contra el SPAM

En muchos países se ha abogado por una solución legislativa nacional contra el SPAM. La Directiva de la Unión Europea 2002/58/EC, del 12 de julio de 2002 ya hace referencia (en su artículo 40) a la necesidad de un consentimiento previo de la persona para el caso de las comunicaciones enviadas para el marketing directo (independientemente del método utilizado,

ya sea fax, correo electrónico o SMS). En España, la Ley de servicios de la sociedad de la información y del comercio electrónico (habitualmente denominada LSSI) prohíbe el envío de comunicaciones comerciales sin consentimiento previo, dejando un tratamiento en mayor profundidad a los códigos de conducta de corporaciones, asociaciones y organizaciones comerciales, profesionales y consumidores.

Desde el 1 de julio de 2006 entró en vigor la nueva redacción de la ley «De la publicidad» en la Federación de Rusia afirma (Naumov, 2006), en la cual apareció una sección 7 que regula la publicidad «difundida por las redes de comunicación electrónica». Esto significa que el SPAM, por primera vez, ha quedado bajo la jurisdicción de la legislación rusa asegura (Kalinin yVlosava, 2007).

Otras legislaciones comunitarias, como la legislación francesa, belga, austriaca, danesa, finlandesa, italiana y del reino unido, contienen medidas similares. Existen leyes más tardías contra el envío no solicitado como puede ser la ley 108-187 (conocida como CAN-SPAM ACT) norteamericana de diciembre de 2003. Esta ley no criminaliza el envío de correo no solicitado, por lo que ha tenido muchos detractores, pero sí criminaliza la utilización de *gateways* para ocultar el remitente y el hecho de no responder a las solicitudes del usuario de excluirse de la lista de correo. En cierta medida, la ley estadounidense debería obligar a los *spammers* a utilizar sus propios servidores con lo que debería ser posible trazar su actividad. La ley contra los mensajes electrónicos no solicitados con fines comerciales (SPAM Act) australiana, aprobada también en diciembre de 2003, exige consentimiento previo, la inclusión de información válida del remitente, y la respuesta obligada del remitente a las solicitudes de eliminación de la lista, imponiendo sanciones muy elevadas en caso de incumplimiento. Además, al igual que la legislación española, la ley de protección de datos australiana (Privacy Act) ilegaliza la recogida indiscriminada de direcciones de correo sin consentimiento del usuario final. Sin embargo, las leyes contra el SPAM ven reducida su eficacia al tratarse de un problema que, como muchos otros

relacionados con redes telemáticas mundiales, tiene un ámbito que supera al nacional. Así, existe un gran número de países sin legislación específica que se pueden convertir en países utilizados por los *spammers* para ocultar sus verdaderas identidades y mantener su negocio en marcha. Hay que tener en cuenta que, según la lista ROKSO de Spamhaus ver (Spamhaus, 2004), un proyecto dedicado a combatir el SPAM en Internet, sólo son doscientas organizaciones o personas individuales las responsables del envío del 80% del SPAM que se genera en Internet. Un efecto inmediato de estas legislaciones ha sido la detención de algunos *spammers* reconocidos o una reducción o abandono de su actividad al considerarse ilegal en el país en el que estos residen. Conviene recalcar que casi un 77% de las organizaciones listadas en el índice ROKSO tienen sede en Estados Unidos. Estas leyes están siendo utilizadas para detener y juzgar a *spammers* por su incumplimiento. Aún siendo Estados Unidos, como se indica en la figura 1.4, la mayor fuente de SPAM confirmado por (Kalinin yVlosava, 2007), no se ha notado aún el efecto del CAN-SPAM Act. Por ello, está por verse un efecto importante de estas legislaciones en el nivel de SPAM que se genera en Internet.

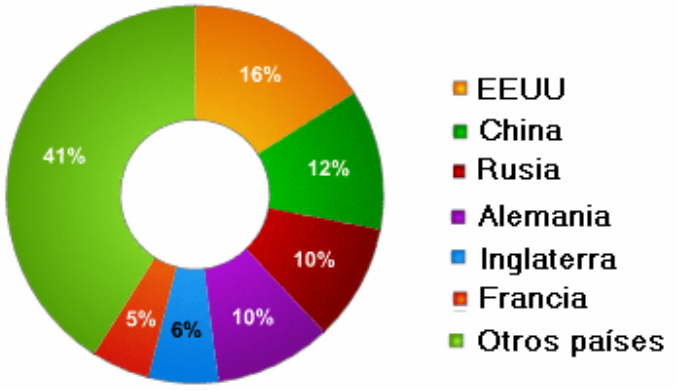


Figura 1.4. Países de donde proviene el SPAM

1.3 Evolución a largo plazo

La evolución del SPAM a largo plazo es, en vista de la evolución de estos

últimos años desalentadora a pesar de los esfuerzos realizados afirma (Fernández, 2005). Por un lado, el volumen de SPAM recibido no hace sino aumentar, por otro lado, se detectan cada vez más las vinculaciones entre grupos dedicados al SPAM y grupos dedicados a otras actividades ilegales, como son la propagación de contenidos maliciosos o la realización de fraudes a través de Internet. Además, el fenómeno del SPAM está traspasando las barreras de Internet y está empezando a llegar a otros servicios. La telefonía móvil sufre también el envío de SPAM a través del servicio de mensajes cortos (*Short Message Service, SMS*), si bien es posible que el costo de dichos mensajes, y la inexistencia de *gateways* gratuitos de mensajería a móviles desde Internet, haya influido en que este fenómeno no haya cobrado aún las mismas proporciones. Otros servicios sufren el azote del SPAM son la mensajería instantánea, los diarios en línea (*blogs*), los foros de noticias vía Web y los servicios de edición colaborativa en línea vía Web (*wikis*). El efecto del SPAM no es muy acusado aún en la mensajería instantánea debido a que, en muchas de estas redes, no es posible enviar mensajes de forma indiscriminada a usuarios que no han «aceptado» al remitente dentro de su grupo de «amigos». Si bien es posible que los *spammers* se aprovechen de vulnerabilidades en estos protocolos (o en equipos controlados remotamente) para llevar a cabo este tipo de envío de mensajes, como ya ha sucedido con virus que han afectado a algunos sistemas de mensajería instantánea. Los otros servicios en línea mencionados han sufrido en mayor o menor medida este tipo de ataques, siendo mayor cuando para estos servicios se han utilizado programas muy difundidos y conocidos, lo que ha facilitado la generación de herramientas automáticas para esta tarea. Este efecto se ha venido reduciéndose en dichos servicios en cuanto se introducía la obligación de los usuarios de registrarse a los mismos. Un último servicio en Internet que ha empezado a sufrir también el SPAM es el servicio telefónico a través de Internet (Voz sobre IP). Aunque la incidencia del SPAM, denominado en estos casos *spit*, ha sido inferior debido al menor número de usuarios. Sin embargo, es posible

que los mecanismos de control sobre el protocolo sean insuficientes para evitar este problema cuando sea utilizado por un mayor volumen de usuarios.

CAPÍTULO 2. HERRAMIENTAS Y ESTRATEGIAS PARA EL CONTROL DEL SPAM

No existen aún soluciones, ni técnicas, ni políticas, para hacer que el SPAM desaparezca o reduzca su incidencia de forma significativa en Internet. Tampoco existe aun desafortunadamente, un consenso en las modificaciones que deben realizarse al sistema de intercambio de correo electrónico (el protocolo SMTP o la infraestructura utilizada en Internet para transmitir correos) para reducir la incidencia de éstos. En la actualidad solo se tienen herramientas y estrategias que permiten minimizar el impacto del SPAM en cierta área delimitada, estando los resultados en base a la correcta elección e implantación de estas herramientas y estrategias.

2.1 ¿Cómo combatir el problema?

El problema del SPAM se debe abordar desde distintos frentes. Desde un punto de vista local, las organizaciones que sufren el problema del SPAM tienen que implementar soluciones para evitar que éste alcance dimensiones poco manejables para sus usuarios y deberán tratar el problema como un riesgo más asociado a la utilización del correo electrónico en Internet. Más adelante se detallarán algunas estrategias que se pueden seguir para intentar abordar el problema del SPAM dentro de una organización (empresa, institución, etc.).

Sin embargo, las soluciones en el ámbito local, deben considerarse soluciones a corto o medio plazo ya que, si bien pueden evitar la aparición, o reducir la incidencia del correo no solicitado, se hace a costa de recursos de

la organización (ya sean personales o materiales) y, a medida que aumente el fenómeno del SPAM en Internet, también lo harán los recursos que este fenómeno consume, ver (Fernández, 2005).

Las soluciones en el ámbito global para este problema son más complejas, por un lado, se debe llevar a cabo cambios en los sistemas de intercambio de correo electrónico para intentar reducir la incidencia del SPAM o hacer más fácil su detección y eliminación. Por otro, se deben abordar cambios en la seguridad de los sistemas informáticos (tanto en los sistemas finales de los usuarios o de las empresas como en los sistemas de los proveedores de acceso a Internet) para hacer más difícil que los sistemas informáticos sean utilizados para estos envíos indiscriminados y para que, en caso de detectarse un envío de correo masivo, se pueda determinar su fuente con exactitud y atajar el problema. Esto, claro está, sin perder de vista que el problema del SPAM, hoy en día asociado al correo electrónico, puede extenderse con facilidad a otras tecnologías y que puede ser necesario, por tanto, adaptar las soluciones adoptadas para este servicio a otras tecnologías en Internet o diseñar las nuevas tecnologías teniendo presente esta amenaza.

2.2 Asegurar la identidad del remitente

Un número muy elevado del correo no solicitado enviado hoy en día utiliza direcciones de correo inexistentes. En algunos casos, se utilizan direcciones de correo válidas (extraídas de las mismas bases de datos utilizadas para el envío) pero que nada tienen que ver con el *spammer* haciendo muy difícil determinar el origen el correo no solicitado.

Por esto los usuarios normales, no acostumbrados al problema, responden con sus quejas al usuario original que, o bien no existe, o nada tiene que ver con el *spammer*. Generando intercambios de correos inútiles que o serán devueltos (porque el remitente no existe) o serán respondidos con sorpresa por parte del remitente.

La única forma efectiva de identificar la fuente del SPAM es a través de un análisis detallado de las cabeceras del correo recibido. Algunas de estas cabeceras, generadas por sistemas intermedios que el *spammer* no controla, ayudan a determinar el sistema real que ha originado el envío (su dirección IP). Sin embargo, este sistema no está, en muchos casos, relacionado con el *spammer* sino que está siendo utilizado por éste para el envío del correo, puede ser un ordenador doméstico, o un servidor proxy o un gateway de correo mal configurada. Lo único que se puede hacer en estos casos es reportar el problema al responsable del sistema (o con el responsable administrativo del rango de direcciones IP en las que está ubicado, generalmente el proveedor de acceso) y esperar a que éste lo resuelva.

2.3 Certificación digital del remitente

Sólo es posible garantizar con seguridad la corrección del remitente (el valor del campo From: en un mensaje) cuando el correo ha sido firmado, utilizando para ello certificados públicos digitales con una cadena de confianza suficiente. Ésta puede ser una práctica útil dentro de una organización, o en el ámbito estatal (si los ciudadanos de un país tienen certificados digitales y saben utilizarlos), y es ya una realidad para los usuarios de correo que hacen uso de certificados PGP ó GPG de manera informal. Sin embargo, la extensión de esta práctica a nivel global es difícil, por la propia dificultad de extender sistemas de criptografía de clave pública a nivel mundial. Cabe recordar que aún no existen autoridades de certificación a ese nivel y es posible que no existan en mucho tiempo dejando de lado (pero no olvidando) la posibilidad de la certificación del remitente con mecanismos criptográficos, no existe un mecanismo para asegurar que el remitente del correo es quien dice ser, pero sí se están desarrollando mecanismos para que se pueda determinar qué *gateway* de correo son las que deberían enviar mensajes asociados a un determinado dominio de correo. Si bien aún no existe un estándar concreto el objetivo de estos sistemas es que los propietarios de un dominio concreto puedan decir: «éstos son los servidores de correo que

están autorizados a enviar correo diciendo que viene de mi dominio». En este caso se trata de certificar el dominio en el From del «sobre» de un correo electrónico (como se define en el RFC2821), no de verificar el remitente indicado en el cuerpo del mensaje, que es la dirección de correo electrónico origen del mismo (definido en el RFC2822). Han surgido múltiples iniciativas compitiendo entre sí, y la gran mayoría de ellas usa el *DNS* para que los responsables de los dominios almacenen información indicativa de los servidores de correo reconocidos para el dominio.

Básicamente, un servidor de correo que reciba una conexión de otro sistema, contrastaría la dirección IP de éste con las direcciones IP publicadas por los administradores del dominio en la información del DNS. Si esta dirección se encuentra dentro de las direcciones o redes indicadas, se permitiría el envío de correo, en caso contrario, se rechazaría.

La solución técnica que parece haber cobrado más fuerza es el convenio de remitentes (*Sender Policy Framework, SPF*) ver (Mehnle, 2006) y (Cea, 2007), una solución para la que ya existen implementaciones para los programas de transporte de correo más utilizados (Courier, Exim, Microsoft Exchange, Postfix, Qmail, y Sendmail). La propuesta inicial promulgada por Microsoft (Caller-ID) ha terminado fusionándose con SPF para constituir una nueva propuesta (Sender-ID) que intenta ofrecer autenticidad del remitente final (no del sistema que envía el correo), según (Microsoft, 2007). Casualmente, el afán de Microsoft por patentar estas técnicas ha llevado a que el grupo de trabajo de la Internet Engineering Task Force (IETF) que las desarrollaba MARID ver (IETF, 2004), tuviera que disolverse en agosto de 2004 .

Aunque aún está por ver el grado de implementación de esta última propuesta, SPF ya está siendo utilizada por un elevado número de dominios. Este tipo de soluciones no tiene como objetivo solucionar el problema del SPAM, pero sí pueden ayudar a determinar si el dominio de un correo ha sido falseado (porque lo remite un *gateway* no reconocido por los administradores del dominio), algo que puede ser indicativo del intento de envío de SPAM. En el caso de que los

spammers utilicen también ellos mismos este mecanismo, como ya está sucediendo con SPF, no se podría utilizar la comprobación para identificar SPAM. Sin embargo, esto permitiría asociar el correo no solicitado a una serie de dominios concretos, y así facilitar la capacidad de reportar el incidente y de intervenir en contra del responsable administrativo del dominio, los servidores de correo de éste o el proveedor de acceso a Internet que le proporciona la conectividad. Además, de llegar el caso en que todos (*spammers* y remitentes legítimos) utilizarán registros SPF, se podrían entonces introducir esquemas de reputación de dominios de forma que, por ejemplo, un dominio reconocido y suficientemente acreditado dispusiera de mayor credibilidad que un dominio recién creado o de un dominio que se sabe está siendo utilizado para enviar correo no solicitado.

2.4 Eliminación de los *gateways* de reenvío

Los mecanismos de envío de mensajes basado en almacenamiento y reenvío (*store & forward*) dan pie a la utilización de *gateways* intermedios o *relays* que pueden utilizar tanto clientes como servidores para enviar correo dirigido a otros servidores distintos. Supuestamente estos *gateways* han de incluir identificación del servidor previo, pero en algunos casos debido a una mala configuración, esto no es así. La mala utilización de estos *gateways* de correo por parte de los *spammers* hace que baste tener un sistema de correo conectado a Internet mal configurado, es decir, que permita el reenvío de correo desde cualquier sistema en Internet, para que una organización se incluya en algunas de las listas negras utilizadas para el control de SPAM algunas de las denominadas (*Real-Time Block List, RBL*), ver (Wikipedia, 2007b). De hecho, en la actualidad, la gran mayoría de las conexiones de envío de correo se realizan contra los sistemas encargados de gestionar un dominio (identificados por los registros *MX*, de Mail Exchange, en el *DNS*) sin utilizar *gateways* de reenvío.

2.5 Estrategias locales contra el SPAM

Una organización que quiera evitar los problemas que el correo no solicitado genera en su propia infraestructura y en sus usuarios, debe desarrollar soluciones

con distintos enfoques. Por un lado se pueden plantear soluciones técnicas al problema y por otro, también son necesarias soluciones de tipo organizativo y social. En cualquier caso, las soluciones reducirán el riesgo asociado con una conexión a Internet, pero no lo eliminarán por completo. Realmente, la única solución que puede evitar que el problema del SPAM desaparezca de forma irremisible es el no intercambiar correo electrónico con Internet, una solución algo drástica y, en muchos casos, inviable. Así pues, habrá que asumir que cualquier solución introducida podrá resolver el problema, en mayor o menor medida, pero no lo resolverá por completo. Generalmente una combinación adecuada de distintas soluciones puede ayudar a reducir el problema en mayor medida que una solución en exclusiva, por muy bien que ésta se lleve a cabo. Algunas soluciones organizativas según (Fernández, 2005) serían:

- La definición de una política corporativa en cuanto a la utilización del correo electrónico en la organización, que especifique tanto las actividades permitidas como las no permitidas en relación con el uso del servicio de correo.
- La formación del personal técnico para que éstos realicen una correcta configuración y supervisión de los sistemas de correo, de forma que sean capaces de reaccionar ante un incidente asociado al envío masivo de correo no solicitado.
- Proporcionar información a los usuarios sobre la política de uso definida por la organización, los problemas del correo no solicitado, los mecanismos para evitar el problema, así como las actuaciones recomendadas en el caso de detectar un problema relacionado con el correo no solicitado.

Las soluciones técnicas al SPAM son muy variadas y podrán implementarse unas u otras en función de los requisitos de la organización. Es decir, no adaptará la misma solución un proveedor de acceso a Internet que una organización. El primero debe ofrecer un sistema de filtrado de correo como medida de protección a un elevado número de usuarios dispares, el segundo debe introducir un sistema de filtrado que implemente las políticas de uso internas, obligando su

cumplimiento. En general suelen consistir en la instalación de sistemas de filtro de SPAM (mediante sistemas específicos de filtrado) y en la instalación de sistemas para reducir el impacto del SPAM a través del control de las comunicaciones a nivel de red de la organización (implementando sistemas de gestión de ancho de banda).

2.5.1 Sistema de filtrado

Los sistemas de filtrado disponibles en la actualidad se pueden dividir en función de su ubicación y en función de la tecnología que utilizan. En la mayor parte de los casos es recomendable combinar los distintos sistemas de filtrado disponibles, ya que son complementarios entre sí.

2.5.2 Ubicación del sistema de filtrado

En función de su ubicación se puede hablar, al igual que en el caso de tecnologías de antivirus, de filtrado en el *gateway* de la organización, en los agentes de transporte de correo (*Mail Transport Agent*, MTA), o de filtrado en el cliente de correo o en los agentes de usuario de correo (*Mail User Agent*, MUA). Los sistemas de filtrado de *gateways* tienden por tener que aplicarse a un conjunto mayor de usuarios, a ser menos adaptables y a tener políticas de filtrado relativamente permisivas, mas detallado, no se pueden permitir falsos positivos, mensajes reconocidos como SPAM cuando en realidad no lo son. En este último caso el usuario deseará recuperar el correo y los mecanismos para ello pueden no estar disponibles o ser de utilización compleja, puede ser necesario proporcionar acceso al *gateway*. Esto obliga a que los mensajes descartados en el *gateway* tengan que ser revisados manualmente por los administradores en caso de que se desee su recuperación, pero es en el *gateways* donde una organización es capaz de aplicar aquellos filtros que más se adecuen a la política de uso que haya definido. Estos filtros podrán estar instalados como primera punto de entrada de correo a la organización directamente expuesta a Internet o tras el *gateway* de primer nivel de la organización, que realizaría un almacenamiento temporal del correo entrante, como se observa en la figura 2.1.

En el caso de que exista un *gateway* con capacidad de detección y eliminación de

virus, su instalación podrá ser anterior o posterior a la misma.

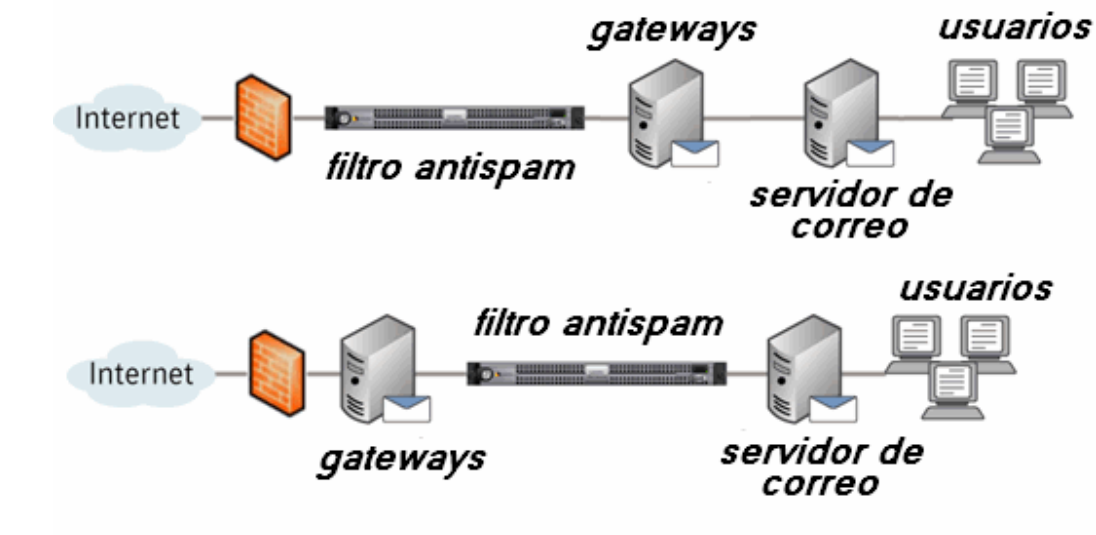


Figura 2.1 Ubicación del sistema de filtrado

No existe realmente una configuración óptima ya que la instalación dependerá del nivel de virus y SPAM que reciba la organización y de los recursos que consuman las soluciones utilizadas, antivirus y anti SPAM. En general, los *gateways* de correo entrante según (Fernández, 2005) podrían:

- Rechazar el correo y generar un rebote (*bounce*) al remitente cuando se detecta que es SPAM. Si bien esto era aceptado hace algunos años ya no lo es pues, como se ha comentado previamente, la gran mayoría de las direcciones origen son falseadas rutinariamente y, de enviarse un rebote, se enviaría a un tercero que en realidad no ha enviado el correo.
- Aceptar el correo e introducirlo en una cuarentena cuando se determina que es basura. La cuarentena se convierte en un elemento adicional a gestionar, ya que los usuarios querrán extraer de la misma correos legítimos mal identificados.
- Aceptar el correo, marcarlo como correo no solicitado (añadiendo cabeceras específicas o cambiando alguna cabecera, habitualmente el asunto del mensaje) y tramitarlo, dejando al cliente de correo que decida qué hacer con él.

- Rechazar el correo dentro de la comunicación SMTP (generando errores 5xx o 4xx) cuando se determina que es basura basándose en el análisis dado por las reglas o filtros implementados.

Esta última forma evita el SPAM generado por sistemas zombi ver (Marcelo.ar, 2005) ya que éstos habitualmente no reintentan el envío, algo que sí harán los *gateways* de correo legítimos bien configurados. También se puede implementar una solución denominada *greylisting*, ver (Harris, 2004) y (Lundgren, 2004) en la que se bloquean temporalmente los sistemas no conocidos, anotando su acceso en una base de datos. Posteriormente, si se realiza una nueva conexión del mismo sistema, se acepta el correo si el remitente y receptor coinciden con el intento de envío anterior. Esto reduce el correo basura a costa de introducir un retardo en el correo recibido de servidores desconocidos, ya que se hace obligatorio un reenvío para éstos. Como ventaja adicional, un sistema no podrá introducir correo si falsean (utiliza de forma aleatoria) las direcciones de remitente o destinatario.

Los filtros de SPAM, especialmente cuando se basan en filtrado de contenidos, análisis de mensaje suponen un importante esfuerzo computacional, con lo que su tasa de entrada o salida de mensajes será inferior a la soluciones de *gateway* de correo tradicionales. Así, puede ser recomendable un *gateway* previo que almacene el correo entrante o saliente. De hecho, éste es el diseño que algunos fabricantes de productos anti SPAM han escogido para poder garantizar el almacenamiento de correo en situaciones de alta carga. No obstante, en este caso, no será posible rechazar el correo antes de que éste entre en la organización, teniendo que descartarse, por tanto, la última de las posibilidades para el tratamiento del correo basura descritas anteriormente.

2.5.3 Tecnologías de filtrado

En función de la tecnología se distinguen los sistemas de filtrado en dos grupos afirma (Fernández, 2005):

- Sistemas de filtrado heurísticos. Implementan un conjunto de reglas específicas para detectar el SPAM.

- Sistemas de filtrado adaptativos. Basados en aprendizaje estadístico y generalmente implementados con filtros bayesianos.

La tecnología de filtrado basada en heurísticos analiza los correos, e intenta determinar si es o no SPAM en función de si cumple o no unas características predefinidas. La figura 2.2 muestra el diagrama de esta tecnología de filtrado.

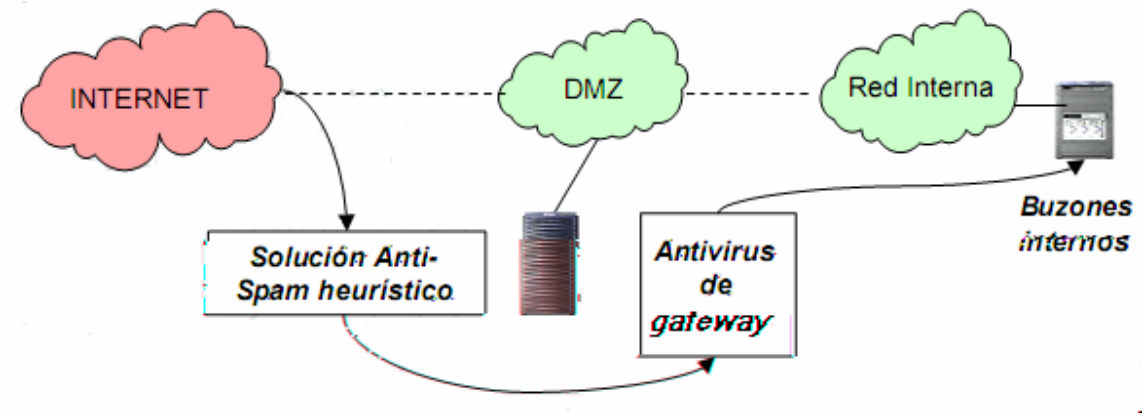


Figura 2.2 Instalación de una solución anti SPAM heurística en el *gateway*.

Algunas de estas comprobaciones son aplicables a un funcionamiento en el *gateway* y otras pueden utilizarse indistintamente de la ubicación. Algunos de los elementos habitualmente utilizados en este análisis son: palabras clave en el cuerpo o cabecera, composición del mensaje (como se utiliza el lenguaje HTML), origen específico del correo (direcciones IP que lo han generado y comprobaciones con listas negras, RBLs o DNSBLs), configuración asociada al DNS del sistema remoto que envía el correo, comprobaciones de los servidores de correo asociados a un dominio (utilizando SPF ó Sender-ID), etc.

Generalmente la tasa de falsos positivos (correos rechazados que son legítimos) dependerá de cada chequeo realizado, pudiéndose producir bloqueos de correos deseados. Además, los spammers tienen como objetivo encontrar mecanismos para esquivar estas reglas y modifican constantemente los contenidos de sus correos, lo que hace necesario revisar las reglas utilizadas e introducir nuevas reglas a medida que los contenidos varían y éstas dejan de ser válidas.

El aprendizaje estadístico, por otro lado, se basa en la generación automática de reglas mediante un sistema al que se le «enseña» qué es SPAM y qué no lo es,

de forma que este genera reglas para detectarlo de forma automática. La figura 2.3 muestra un diagrama de esta tecnología.

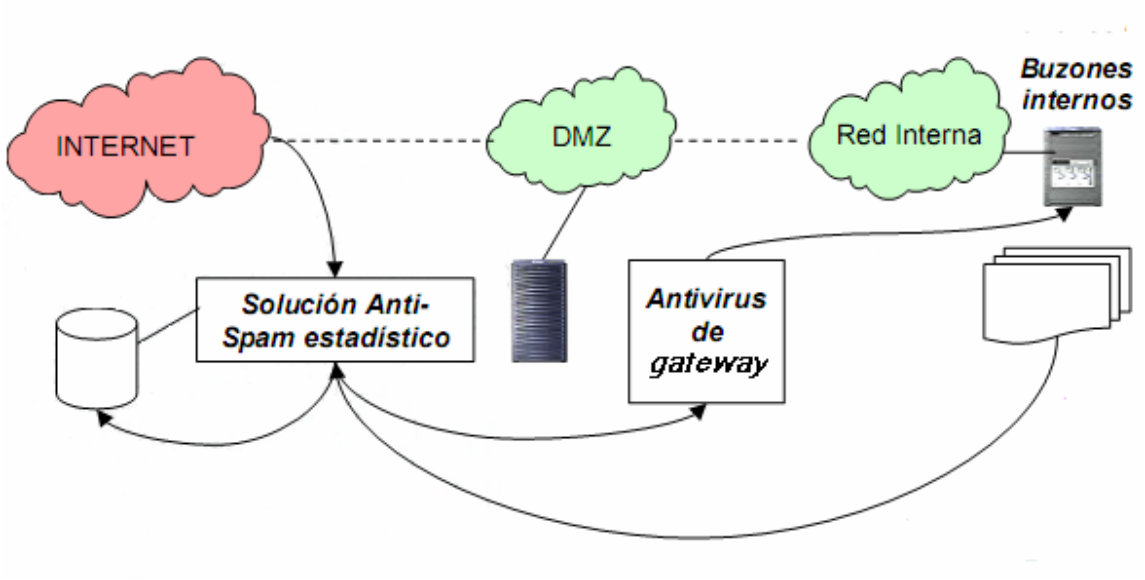


Figura 2.3 Instalación de una solución anti SPAM estadística en el *gateway*.

Las reglas se generan analizando el mensaje completo (cabeceras y cuerpo) y sirven para crear clasificaciones (grupos) de correo no solicitado de forma que sea posible determinar, con un cierto grado de confianza, si un correo nuevo luego de efectuado el análisis pueda agruparse como no deseado. Esta tecnología es útil para extraer reglas no evidentes del SPAM y es capaz de adaptarse a los nuevos tipos de correo no solicitado que van surgiendo, siendo necesario, sin embargo, un aprendizaje permanente.

Estos sistemas de aprendizaje se implementan habitualmente en los agentes de correo del usuario final, dado que el usuario puede ir «enseñando» al sistema marcando el correo que recibe como SPAM. También es posible implementarlos en sistemas de *gateway*, pero para ello es necesario utilizar direcciones falsas que se publican con el objeto de que las utilicen los *spammers* y cuyos buzones alimentan constantemente a la herramienta de aprendizaje.

2.5.4 Gestión del ancho de banda

Es importante recalcar que, si bien una solución de filtrado puede ayudar

mucho a reducir el problema del SPAM, el simple hecho de recibir correo (para posteriormente descartarlo) supone un consumo importante de recursos para la organización. En el caso de que tenga una efectividad del 100%, se estarán gastando además, recursos que dependen directamente del volumen de correo manejado.

En el caso de sufrir una oleada de intentos de envío de SPAM, no sólo sería necesario ser capaz de absorberlo (disco suficiente en los *gateways* para almacenarlo) sino que sería necesario tratarlo con cuidado, para que éste no afecte al correo legítimo. Al incrementarse la entrada de correo no solicitado y su tamaño se incrementan también las necesidades de disco y CPU de los sistemas que lo tratan de forma equivalente. Un ejemplo muy claro de este problema es la situación de muchos proveedores de acceso a Internet que están sufriendo, día a día, los embiste de los *spammers* y tienen que revisar e incrementar el número de sistemas que dedican para tratar el correo de sus usuarios y ser capaz de absorber, al mismo tiempo, el correo no solicitado que les llega. La implementación de sistemas de filtrado puede aliviar el problema a corto plazo, pero a largo plazo es necesario implementar otras soluciones para controlar el volumen del correo. Por esto es necesario complementar los sistemas de filtrado con sistemas de gestión del ancho de banda, que permitan, por un lado, controlar y ralentizar la entrada de correo basura y por otro, garantizar que otros servicios que hagan uso de la misma línea de conexión a Internet, no se vean afectados en caso de sufrir un envío masivo de correo; la figura 2.4 indica como sería la implementación de esta solución. Al mismo tiempo debe garantizarse que los intercambios de correo con servidores de correo de confianza (servidores de correo corporativos en oficinas remotos, clientes prioritarios o proveedores de Internet locales) es priorizado frente al intercambio con servidores de correo desconocidos.

Para ello se pueden utilizar sistemas genéricos de gestión de ancho de banda existentes en el mercado y capaces de gestionar múltiples protocolos de Internet. Estos sistemas pueden utilizarse para implementar criterios de gestión de ancho de banda para los servicios que una organización utiliza o

provee en Internet aplicando su política corporativa, si la hubiera.

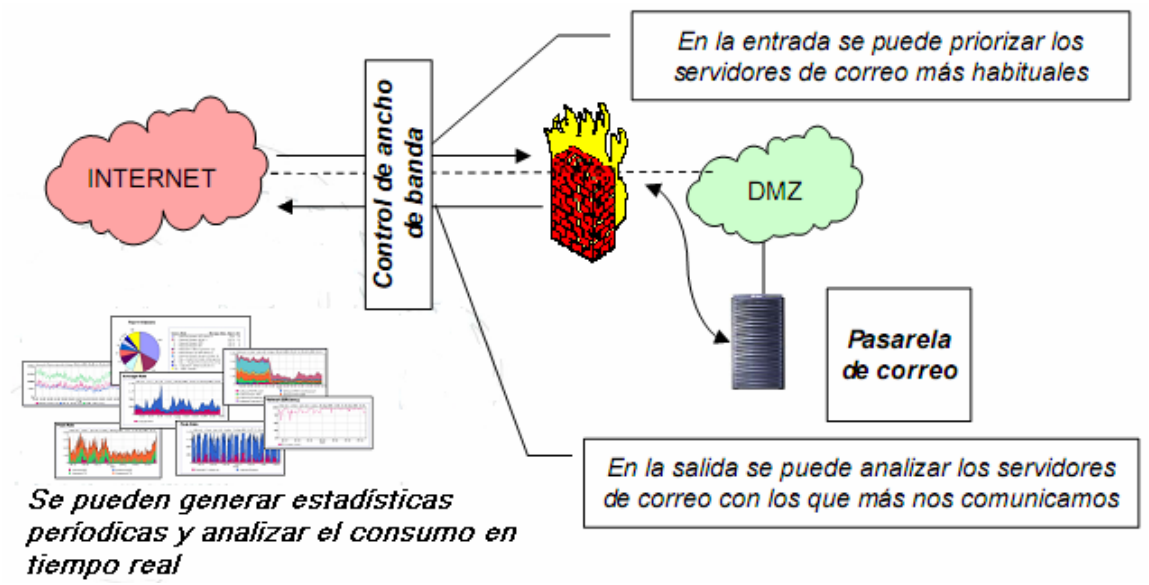


Figura 2.4. Solución para la gestión de ancho de banda.

También es posible utilizar programas en servidores de correo que sean capaces de imponer un cierto control del ancho de banda. Estos servidores de correo modificados realizan también comprobaciones del correo entrante que recibe (origen de la conexión, identificación del remitente en el «sobre» del correo, etc.) para intentar determinar si puede tratarse de un correo no solicitado, en caso de serlo, ralentizan la conexión del sistema remoto, por ejemplo, respondiendo a las órdenes que éste envía en la conversación SMTP más despacio de lo habitual.

2.6 Las empresas y sus productos anti SPAM

Hasta la fecha, diversos son los productos desarrollados por grandes y pequeñas empresas para combatir el SPAM. Aunque el problema no ha desaparecido completamente si se evidencian mejores rendimientos en la detención y aún más cuando se utiliza la combinación de dos o más productos. Como la última generación de correos no deseados incorporan tácticas sofisticadas como aleatoriedad extrema, ocultamiento del origen, y evasión de filtros mediante el uso de HTML, se obliga al constante desarrollo de los filtros para mantener inmunidad y robustez frente al SPAM.

Entre las empresas líderes en esta tarea se encuentran SPAMfighter, Symantec y McAfee con toda una gran gama de productos anti SPAM los cuales se encuentran a disposición en los sitios correspondientes, esto no quiere de decir que son los únicos con prestigios en el mercado, pues tenemos al *spamassassin* creado por Justin Mason que tiene muy buen rendimiento en la detención de los correos no deseados, no pertenece a ninguna de estas empresa y además es un *software* libre.

Entre la diversidad de productos existentes encontramos los destinados a usuarios *MUA* como el Mozilla, el Outlook y los destinados a servidores *MTA* como el *spamassassin*, el SpamFighter Exchange Module y el McAfee GroupShield, pero en realidad los destinados a servidores son mejores por ser más efectivos. Esta efectividad radica en la centralización de las actividades como la creación, instalación y mantenimiento de reglas que posibilite el funcionamiento del producto, alejando al usuario de estas acciones que resultan engorrosas y requieren de experiencia en el tema.

2.7 Filtrado anti SPAM en Servidores Microsoft Exchange 2007

Para el filtrado del SPAM en servidores basados sobre Windows como Microsoft Exchange 2007 tenemos diferentes *software*:

- Symantec Mail Security 8200
- SpamFighter Exchange Module
- SPAMfighter Hosted SPAM Filter
- SPAMfighter SMTP Anti SPAM Server
- McAfee GroupShield

Uno de los más aceptados es el módulo SPAMfighter Exchange (SEM), porque es la solución anti SPAM más fácil de usar para servidores de Microsoft Exchange en pequeñas y medianas empresas, creado por SPAMfighter, compañía líder en Europa en tecnología anti SPAM y actualmente protege a más de 2.500 empresas en todo el mundo

Hay 2 versiones básicas de este programa que se muestran y comparan a continuación.

Tabla 2.1. Comparación entre la dos versiones del módulo SPAMfighter Exchange

Características	**SPAMfighter Standard**	**SPAMfighter Pro**
Protección automática de «correo real»	✓	✓
Protección de múltiples cuentas de correo	✓	✓
Lista negra de dominios y direcciones	✓	✓
Denuncia de SPAM con sólo un clic	✓	✓
Gestión automática de la Lista Blanca	✓	✓
Protege contra el "*phishing*", robo de identidad y otras estafas electrónicas	✓	✓
Disponible en 18 idiomas	✓	✓
Derecho a utilizar SPAMfighter en una empresa/organización	✗	✓
Herramienta exclusiva de filtrado de idioma - Bloquee correos en idiomas que no comprende	✗	✓
Opción de desplazar la barra de herramientas en Microsoft Outlook	✗	✓
Servicio de Soporte de primera calidad - Respuesta en menos de 24 horas, consulta gratuita por teléfono	✗	✓
Entradas ilimitadas en las Listas Negra y Blanca	✗	✓
Opción de eliminar la firma SPAMfighter de sus correos	✗	✓
Cliente de correo libre de publicidad	✗	✓
Precio (incluye todas las actualizaciones y soporte técnico)	**Gratis**	US$29

Este filtro anti SPAM es muy simple de usar. El SPAM es automáticamente trasladado a la carpeta de SPAM, pues conforme llega el correo, el servidor

Exchange lo envía a sus destinatarios, el SEM realiza una firma encriptada cifrada única para cada mensaje y la envía al servidor de SPAMfighter para ser evaluada. Si el servidor SPAMfighter determina que un mensaje es SPAM, se lo notifica al SEM y es trasladado a la carpeta de SPAM del usuario.

Este módulo está alimentado por 3.239.760 SPAMfighters de 215 países o áreas y cuando un número suficiente de SPAMfighters denuncia el mismo correo SPAM, este es inmediatamente eliminado de todos los SPAMfighters y usuarios del SEM. Esto implica una protección instantánea sin configuración o mantenimiento perfecta para pequeñas y medianas empresas que trabajan con poco o ningún personal técnico.

El Módulo SPAMfighter Exchange se integra perfectamente con Microsoft Exchange Server, es muy rápido, permite una administración e instalación fácil y es transparente al usurario aunque proporciona control individual para los usuarios si los administradores lo deciden así. También ofrece informes, de manera que los administradores pueden revisar estadísticas acerca de cuántos mensajes el SEM esta filtrando, cuántos usuarios están disponibles, etc.

Características:

- Protección inmediata contra SPAM y Phishing
- Seguro y libre de mantenimiento
- Privacidad garantizada, sus correos nunca abandonarán la red
- Filtro de SPAM opcional por usuario
- Cero Posibilidades de indisponibilidad debido a SEM
- Procesado de SPAM y estadísticas de bloqueo
- Recomprobación del buzón del Servidor cada 60 minutos
- Funciona automáticamente sin configuración

- Fácil de usar para los usuarios terminales, el SPAM es automáticamente trasladado a la carpeta de SPAM

Requisitos para la instalación:

Sistema Operativo: Microsoft Windows Server 2000 o más nuevo

Servidor de correo: Microsoft Exchange Server 2000. 2003 o 2007

Software adicional: Internet Information Services

Microsoft .NET Framework (1.1 SP1 o 2.0)

Microsoft Data Access Components 2.7

Memoria Ram: 256 MB mínimo

Espacio en disco: 10MB

Procesador: Velocidad de la CPU 1000 MHz

2.8 Exim4 Agente de transporte de Correos para servidores Unix

Exim (*Experimental Internet Mailer*) es un agente de transporte de correo desarrollado por la Universidad de Cambridge ver (García, 2001). El proyecto comenzó en el año 1995 y su objetivo principal era crear un MTA para ser usado en la universidad. Sin embargo, desde el comienzo, el sistema fue utilizado por otros organismos e instituciones que fueron accediendo al producto en base a distribución personal ya que las primeras versiones nunca fueron anunciadas.

Este *MTA* fue creado en base al código fuente de Smail 3 (de Ron Karr) y puede ser utilizado en la mayoría de los sistemas UNIX (entre ellos Linux). El servidor puede compilarse en sistemas operativos Windows pero se recomienda que sea utilizado en producción sobre la familia UNIX. Se distribuye sin costo bajo la licencia GPL de la FSF por lo que es un *software* libre que satisface perfectamente las necesidades de un servidor SMTP normal.

A partir de la versión 4 lanzada en Febrero del año 2002, se introducen cambios importantes respecto a la versión 3, destacándose las políticas de control de correo entrante denominadas listas de control de acceso (*Access Control Lists,*

ACL) que incorpora quizás el mecanismo más sofisticado y más flexible para el filtrado del SMTP en tiempo real.

Las ACL se pueden utilizar para evaluar si aceptar o rechazar los pasos de la transacción del mensaje entrante, el inicio de una conexión desde un host remoto o algunos comandos SMTP como el HELO/EHLO, MAIL FROM y el RCPT TO explicadas en (De CocK y otros., 2006).

Las listas de control de acceso consisten en una serie de declaraciones o de reglas que se evalúan en orden, hasta que se tome una acción definitiva. Cada regla comienza con el verbo de la acción, por ejemplo acepta (*accept*), advierte (*warn*), requiere (*require*), difiere (*defer*), o niega (*deny*), seguido por una lista de condiciones, de opciones, y de otros ajustes referente a la misma afirma (De CocK y otros., 2006).

Un ejemplo simple de una regla es la siguiente aplicada a RCPT TO.

```
deny
  message  = relay not permitted
  !hosts   = +relay_from_hosts
  !domains = +local_domains : +relay_to_domains
  delay    = 1m
```

Esta regla no hace más que rechazar el RCPT TO, sino fue entregado por un host de la lista de host "+relay_from_hosts" o un dominio de la lista de dominio "+local_domains" o "+relay_to_domains". Considerando que antes de publicar la respuesta SMTP “550”, el servidor esperará un minuto.

Es preciso destacar que para evaluar una ACL en particular en una etapa dada de la transacción del mensaje, necesitas señalar uno de los controles de la política de Exim a esa ACL.
Por ejemplo, utilizar el acl_rcpt_to mencionado anteriormente para evaluar el RCPT TO, la sección principal de tu archivo de la configuración de Exim antes de comenzar las palabras claves debe incluir:

```
acl_smtp_rcpt = acl_rcpt_to
```

Para crear una lista completa de tales controles de la política.

La versión actual es la 4.66 y está disponible desde el 2006, cuenta con 88.179 líneas de código fuente y más del 98% del código fuente de la distribución está escrito en lenguaje C.

Exim es utilizado usualmente en conjunto con Clamav (antivirus liberado bajo la licencia GPL) *spamassassin* (proyecto de la Apache Software Foundation para el control de correo SPAM, liberado bajo licencia Apache2) y control de bloqueo mediante listas dinámicas basadas en DNS.

2.8.1 Escritura general del fichero de configuración de Exim

El fichero de configuración se divide en 6 bloques. Cada bloque está separado del siguiente por la palabra ‘‘end’’, excepto el último que no lleva ‘‘end’’. Todos los bloques deben aparecer, si alguno se encuentra vacío tiene que aparecer el ‘‘end’’ de todos modos. Los bloques son los siguientes:

1. **Configuración principal**: aquí van las directivas principales de configuración, las preferencias como el nombre de nuestra máquina, a quién se le hace *relay*, y otros.
2. **Transports**: Cuando se sabe definitivamente cómo y a dónde se va a enviar un determinado mensaje, el *transport* correspondiente es el que se encarga de hacerlo. Cada *transport* tiene un *driver* que indica el tipo de reparto. Ejemplos de *drivers*: ‘‘*appendfile*’’, que concatena el mensaje a un fichero (para folders UNIX normales), ‘‘smtp’’ que hace una conexión a un smtp para enviar.
3. **Directors**: Cuando un mensaje va a una dirección local, se busca un director que sepa qué hacer con ella. Son los que se encargan de buscar en el fichero de aliases, etc. El orden es importante.
4. **Routers**: cuando una dirección no es local, se busca el primer *router* que sea capaz de enviarla. El orden es importante.
5. **Retry**: Aquí se especifica el tiempo que tiene que transcurrir hasta que se considere que un mensaje no se puede enviar.
6. **Reescritura**: Aquí están las reglas de reescritura de cabeceras.

2.9 spamassassin

spamassassin es una herramienta realizada por Justin Mason en el 2001 como una opción para el filtrado de correo SPAM, consiste en un mecanismo heurístico basado en reglas y ponderaciones predefinidas incorporados en un algoritmo bayesiano afirma (Committee, 2007). Es una integración de elementos que combinan varias técnicas para la detección de SPAM. Originalmente registrado por DeerSoft, posteriormente adquirido por Network Associates, actualmente opera bajo la Licencia Apache Software Foundation.

Sus características primarias son:

- Chequeo de la cabeza del mensaje.
- Chequeo del cuerpo del mensaje.
- Filtrado con técnicas Bayesianas.
- Chequeo automático de direcciones en listas blancas y negras.
- Chequeo manual de direcciones en listas blancas y negras.
- Bases de datos de colaboración para la identificación del SPAM (DCC, Pyzor, Vipul's Razor,).
- DNS Blocklists, conocido como RBLs o Realtime Blackhole Lists.

El producto realiza un análisis de cada mensaje, aplicando una serie de reglas predefinidas, las cuales en conjunto son el corazón del sistema de detección.

Aprovecha estas reglas rápidas para identificar el 95% de los mensajes, dado que la gran mayoría de los *spammers* no escriben su propio código y utilizan algún sistema abierto que adiciona encabezados. Es relativamente sencillo identificar cuando se altera un mensaje, destacando que por defecto intenta reforzar sus propias reglas con la filtración Bayesiana, siendo más eficaz con la entrada real del usuario.

Ante el surgimiento de *spamassassin* la reacción de algunos *spammers* fue empezar a utilizar encabezados que simulaban ser legítimos durante el envío masivo de mensajes, por ejemplo desde Outlook Express.

2.9.1 Las primeras reglas de filtrado

El *spamassassin* realiza un análisis en los encabezados o el cuerpo del mensaje para identificar coincidencias de patrones, (por ejemplo: Haga dinero fácil…), búsquedas de resolución de DNS (por ejemplo: sco.com, 127.0.0.1, etc.) o también una verificación (*checksum*) del mensaje. Sin embargo aún en encabezados modificados, un mensaje correcto contiene un identificador válido único.

```
Message-ID: <00e601c4926f$468d5870$....
Message-ID: <008901c48e89$3efc91a0$....
Message-ID: <0b5301c49704$2741bee0$....
Message-ID: <006d01c493f3$c570da60$....
Message-ID: <002801c32fd1$5f6598f0$....
```

Aun Outlook Express incluye en forma automática un identificador para la fecha en el identificador de cada mensaje.


```
Message-ID: <0b5301c49704$2741bee0$....
```

Se genera el identificador del mensaje y se compara, permite incorporar la regla MSGID_OUTLOOK_INVALID la cual detecta un 25% del SPAM.

Las principales características del *spamassasin* son:

Aleatorio: permiten utilizar una plantilla para el envío masivo de mensajes, incluyen mecanismos que toman algún valor del entorno. La regla

PERCENT_RANDOM permite identificar un 17% del SPAM.

Verificación por *Checksum*: Las bases de datos Vipul Razor, Pyzor, DCC Permiten revisar cuando un mismo mensaje se envía a una gran cantidad de sistemas o destinatarios, permite su detección al adicionar un encabezado o firma y compararlo con alguno de estos sistemas que manejan listas válidas (*Whitelists)*. Se puede configurar *spamassassin* para reportar a estos sistemas el envío de SPAM.

Reglas: Se cuenta con una serie de reglas predefinidas que nos permitirán iniciar la ponderación del mensaje, por ejemplo:

- *20_porn.cf* indicadores de encabezados porno
- *20_dnsbl_tests.cf* para las pruebas de Listas Negras DNS
- *20_phrases.cf* identifica frases para ser removido

Otras características a tener en cuenta en esta herramienta es que ninguna regla, por sí sola, puede marcar un mensaje como SPAM y que adiciona la posibilidad de aprender a clasificar el mensaje en base a un grupo de carpetas, en donde el usuario previamente han incluido sus mensajes basura (SPAM) y mensajes válidos (HAM). Esta operación le permite a *spamassassin* «aprender» a identificar cada correo.

Las reglas de alguna forma son el corazón el *spamassassin*. Hay varios sitios en Internet que crean juegos de reglas diariamente. Estos juegos de reglas se agrupan en estructuras que se denominan canales.

El *spamassassin* a través de la utilidad *sa-update* puede sintonizarse a uno o varios de estos canales manteniendo así sus bases de datos y reglas actualizadas para poder mejorar la calidad de la clasificación.

Verdaderos Negativos (HAM).

Son aquellos mensajes en que el usuario y SA están de acuerdo en que no son SPAM. Adiciona el encabezado **X-Spam-Status** con la leyenda **NO** y **X-Spam-Checker-Version** con la versión utilizada por SA.

Verdaderos Positivos (SPAM).

Son aquellos mensajes en que el usuario y SA están de acuerdo en que es SPAM, adiciona los encabezados **X-Spam-Level**, **X-Spam-Status**, y **X-Spam-Flag**. Si se habilita la opción **rewrite_subject** se adiciona en el título del mensaje ********SPAM********.

2.9.2 Instalación

SA se escribió para entornos basados en UNIX que incluyan Perl, de preferencia 5.6.1 o recientes. Se requieren los módulos *ExtUtils::MakeMaker*, *File::Spec*, *Pod::Usage*, *HTML::Parser*, *Sys::Syslog*, *DB_File*, *Digest::SHA1*, y *Net::DNS*. Se pueden consultar tres sitios de referencia *Vipul'sRazor* (http://razor.sourceforge.net), *Pyzor* (http://pyzor.sourceforge.net), y *DCC* (http://www.rhyolite.com/anti-spam/dcc/).

2.10 Para contener la situación

Varios y diversos son los productos disponibles tanto gratis como rentados para hacer frente al problema, ejemplo de ello es el *spamassassin*, las reglas del Exim4, el Outlook entre otros obteniendo una buena inmunidad. Los filtros anti SPAM hasta ahora desarrollados, constituyen una herramienta muy factible. Incluyen nuevas reglas para el filtrado que deben estar bien definidas y configuradas, subrayando la actualización constante de la mismas, pues generalmente una regla de filtrado puede dejar de ser efectiva en tan solo una semana, por la rápida evolución de las técnicas realizadas por *spammers* para burlar los filtros, unas de las características que convierte en un verdadero reto la eliminación radical del problema de Internet. Hoy al menos se puede contener la situación.

CAPÍTULO 3. SITUACIÓN RED UCLV

3.1 Red UCLV

La Universidad Central «Marta Abreu» de Las Villas cuenta con una red con más de 2000 computadoras repartidas por 13 nodos y 14 Sedes Universitarias y un directorio de más de 10000 usuarios, ver distribución en la figura 3.1 que muestra la estructura lógica de la red UCLV.

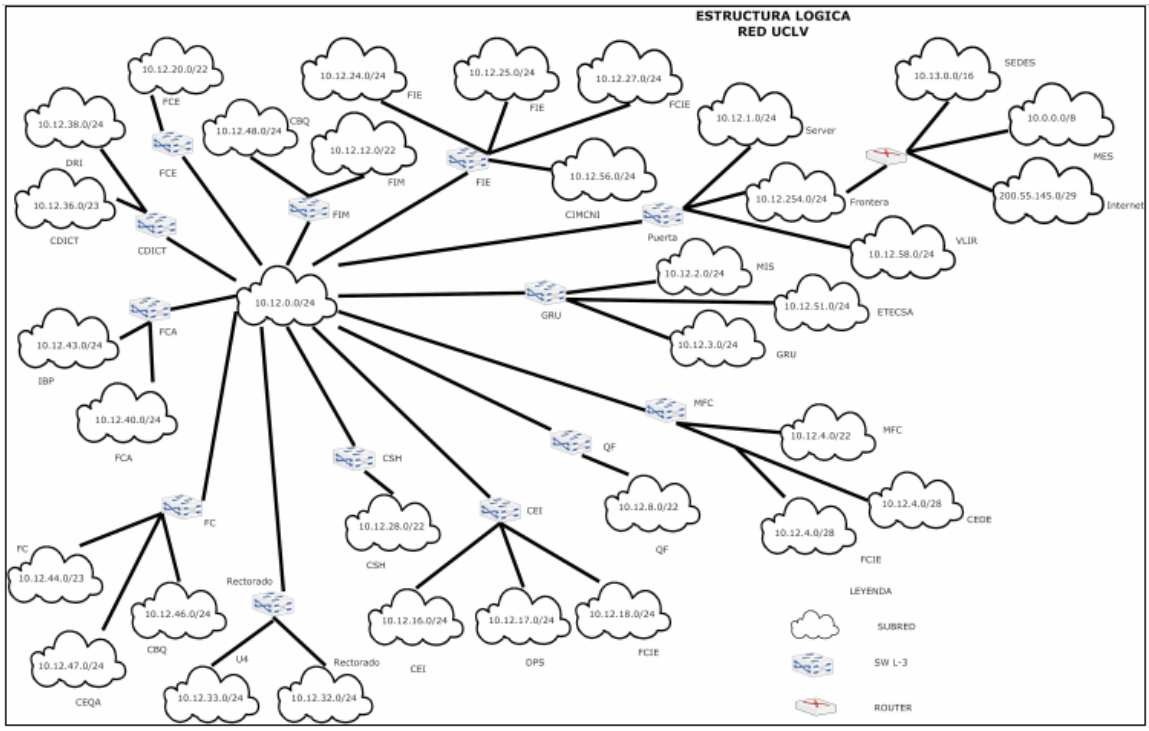


Figura 3.1. Estructura lógica de la red UCLV.

El nodo central está formado por 24 servidores y un mínimo de 2 servidores en cada uno de los nodos que forman el Campus Universitario, afirma (López, 2007), ver figura 3.2

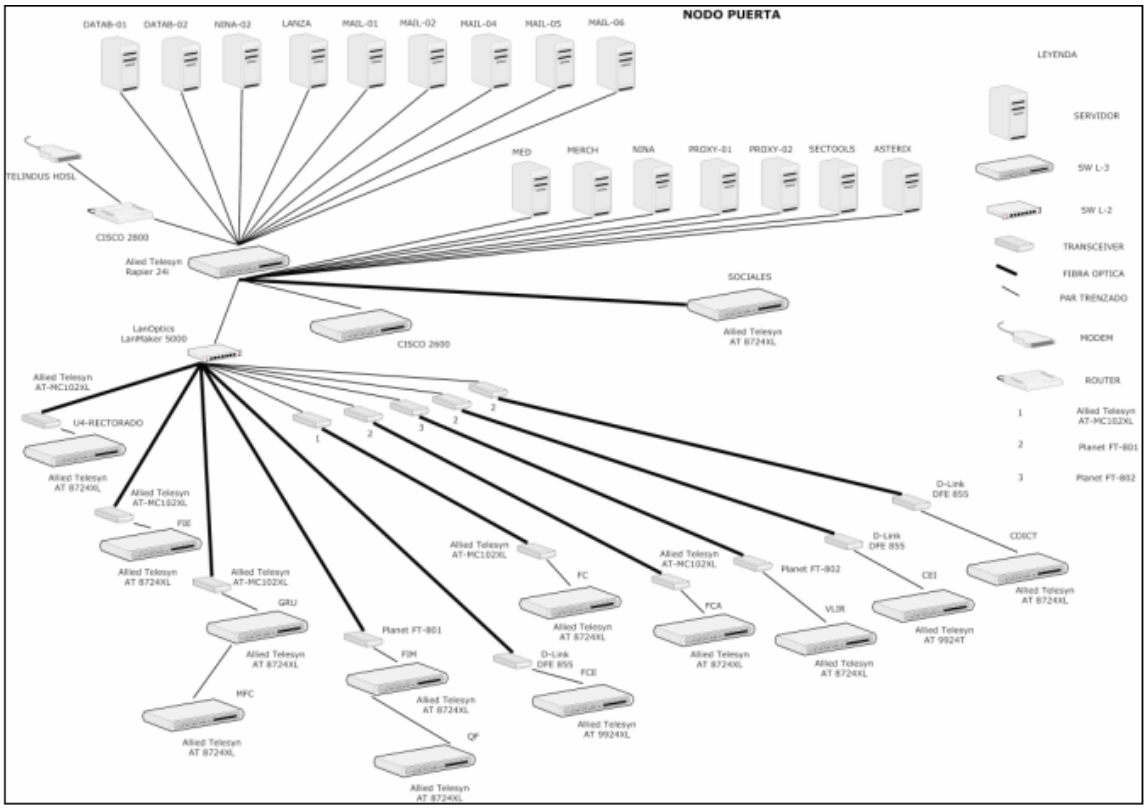


Figura 3.2. Nodo central

El *backbone* de distribución de la universidad está constituido por un tendido de fibra óptica multimodo (62.5/125µm) el cual conecta la totalidad de las facultades y dependencias del campus universitario, además existe un enlace de fibra óptica monomodo (9/125 µm) con la Estación Experimental y la dirección del proyecto VLIR (Casa Belga), ver (López, 2007). Su estructura se basa en una tipología física en estrella con tres niveles jerárquicos. El primer nivel se encuentra en el nodo principal de conmutación de la red ubicado en «La Puerta» donde existe un *patch-panel* que conecta todos pares de hilos de fibra óptica que se difunden por la universidad. El segundo nivel está representado en el Centro de Estudios de la Informática (CEI), en el Edificio Administrativo (U4) y el nodo ubicado en el edificio de Ciencias Sociales y Humanísticas (CSH),

El nodo de conmutación de la red universitaria, descansa sobre un *switch* con estructura modular modelo *LanMaker 5000*. La red universitaria se conecta al exterior utilizando un Modem Telindus a través de una línea arrendada contratada

a ETECSA, en la frontera de la red universitaria se encuentra un *router* Cisco 2800 encargado del encaminamiento entre la Red UCLV y el exterior, el mismo cuenta con dos interfaces conectadas al *switch* Rapier configuradas con el objetivo de independizar el tráfico generado por la red del MES (reduniv) del tráfico generado por el acceso a Internet. Por otra parte, al *switch* Rapier se conectan los servidores ubicados en «La Puerta» y el *switch* LanMaker, quedando conectada de esta forma toda la red de la universidad a través del *backbone*. Además de los elementos antes mencionados, existe un *router* Cisco 2600 para atender el Servicio de Acceso Remoto con capacidad para 8 conexiones telefónicas, este dispositivo está directamente conectado al *switch* Rapier. La figura que se muestra en el Anexo 4 ilustra con mayor claridad la estructura física de la red.

3.2 Situación actual

La UCLV al igual que los demás centros de educación superior de nuestro país es una red diseñada para el apoyo a la docencia y la investigación. Por ello existen un grupo de reglamentaciones para controlar el uso del correo electrónico.

De forma resumida se puede decir que hay dos grupos de usuarios en cuanto a correo electrónico. Uno que tiene entrada y salida solo nacional y otro que puede enviar y recibir correos desde cualquier dominio.

Esta clasificación ahorra un poco de trabajo, pues las reglas se pueden simplificar y evitar que una gran parte de los usuarios de la red no tenga problemas de SPAM. Es bueno tener presente que casi la totalidad del SPAM que llega a la UCLV no tiene dominios .CU (Cuba).

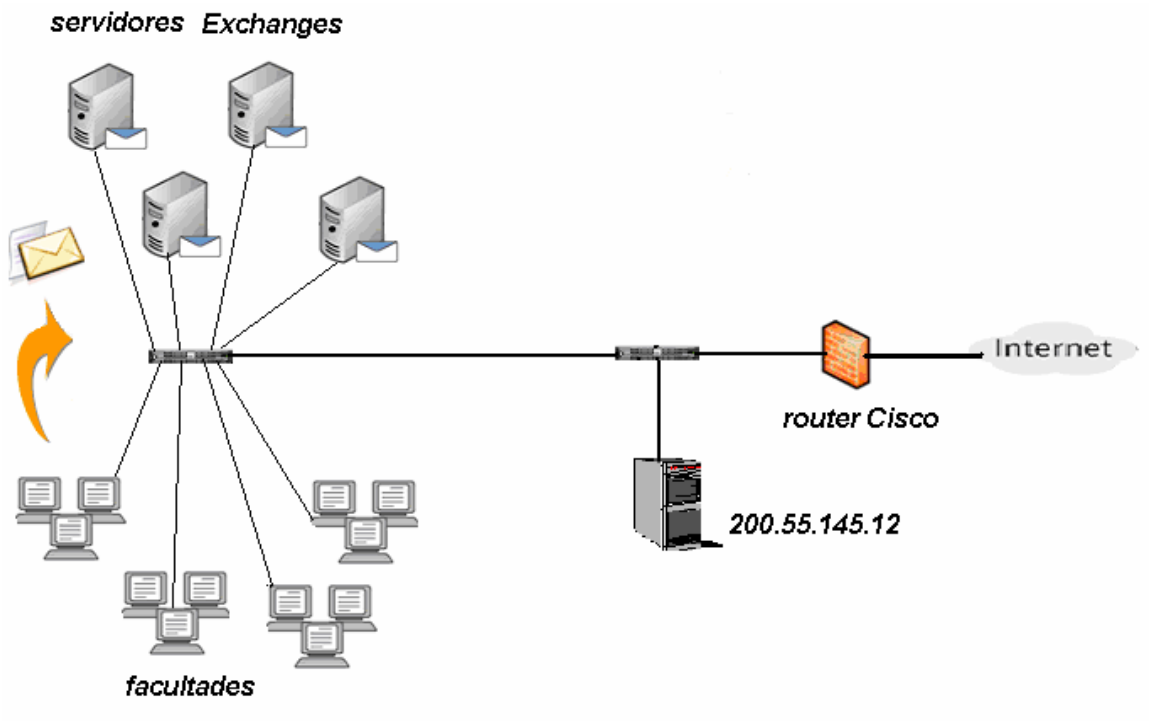
Para el caso de los usuarios que si tienen acceso desde dominios diferentes a .CU se hacen varias validaciones. Por ejemplo se trata de identificar el nombre del servidor desde donde están enviando los correos, se intenta validar si es una dirección válida, y se comprueba que no se esté en listas negras.

Para analizar en detalle la configuración actual y las mejoras implementadas se debe separar el esquema en dos partes según el flujo de correos: la entrada y la salida

Salida de correo.

La salida está controlada por los servidores de Exchange que son los usados para dar la cara a los usuarios de la red y donde están almacenados todos los buzones y los correos. Se usan las listas de *Active Directory* de correo Internacional y correo Nacional para validar quien puede salir y quien no.

Luego los correos van al servidor con dirección 10.12.1.24 para ser entregados a través de Internet, antes de salir, se adiciona al final del correo un pie de firma identificando el servicio de la UCLV.



Como el proceso de salida no está asociado con problemas de SPAM, y en la UCLV no se generan, no hay instalados muchos controles para este problema. Solo se hace un control de la cantidad de destinatarios que lleva un correo y si es un número muy alto se bloquea el correo alertando al administrador del sistema para prevenir la aparición en una lista negra como productores de SPAM.

Entrada de correo.

La entrada de correos es el problema mayor en cuanto a control de SPAM. Como se muestra en la figura 3.1 los correos pueden entrar por dos estaciones de forma

aleatoria, esto se hace para garantizar redundancia en el servicio.

En esos dos servidores se analiza cada uno de los correos que entran. Se bloquean o se aceptan en base a los siguientes datos: IP de origen, nombre del servidor que envía, dirección de la persona que envía y dirección del destinatario.

Si el correo pasa este primer chequeo realizado por las reglas del *MTA* Exim4, deberá ser evaluado para ver si es SPAM o no. Para eso se pasa por el *spamassassin*, que se está configurado para usar las reglas suministradas por saupdates.openprotect.com

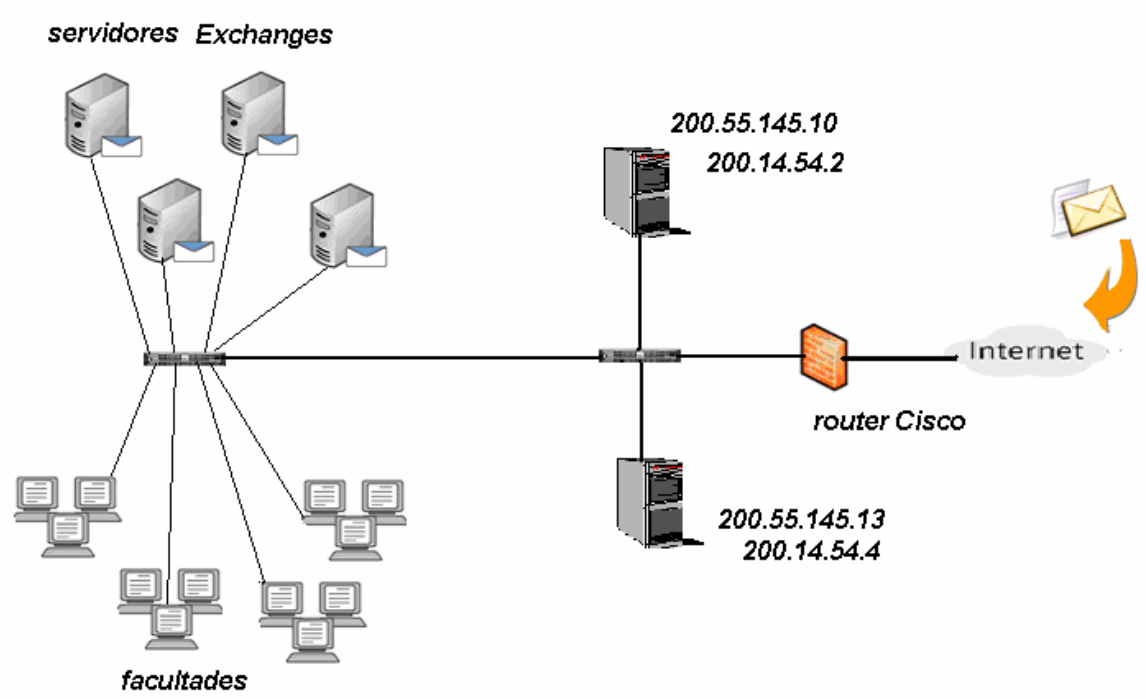


Figura 3.1. Servidores de Correo en la UCLV

Es bueno tener presente que una vez que el correo entró completamente al servidor no tiene mucho sentido descartarlo, pues ya consumió el tiempo y la parte del canal de comunicaciones que estamos cuidando. Es por ello que se hace muy importante tratar de validar el correo antes de que entre completamente, o sea, solo con las líneas HELO, MAIL FROM, RCPT TO.

Si el correo pasa todas estas verificaciones se entrega entonces a los Exchanges para que sea almacenado en buzón del usuario final.

3.3 Configuración de Exim4

La *ACL* más importante del Exim en cuanto a control de entrada es la check_rcpt porque en el momento en que se aplica ya que se tienen los principales datos del correo y no se ha aceptado en el cuerpo que es lo que más tamaño ocupa.

A continuación se explica paso a paso las reglas establecidas en los servers de la UCLV. Estas reglas son bastante genéricas así que pueden ser fácilmente portadas a otros servidores.

Se aceptan todos los mensajes generados localmente. Se usa como método de pruebas. También para los correos generados por WebMail.

```
acl_check_rcpt:
  accept
    hosts = :
```

Se crean cuatro listas, dos para tener nombres y las direcciones IP de los *server* que se aceptarán siempre, pues se consideran seguros. Y dos listas más para lo contrario, o sea, los que nunca se aceptarán.

```
  accept
    domains = +local_domains
    senders = lsearch;/etc/exim4/whitelist_from

  accept
    domains = +local_domains
    hosts   =  lsearch;/etc/exim4/whitelist_ip

  deny
    message = sender envelope address $sender_address is locally blacklisted here
    senders = lsearch;/etc/exim4/blacklist_from

  deny
    message = sender IP address $sender_host_address is locally blacklisted here.
    hosts   = lsearch;/etc/exim4/blacklist_ip
```

Se controlan varias formas erróneas de presentación de máquinas que generalmente son usadas por *spammers*.

```
  deny
```

```
   !hosts = +relay_from_hosts
   condition = ${if match {$sender_helo_name}{\N^[^.:].*[.:][^.:]+$\N} {no}{yes}}
   message = Tu HELO es incorrecto $sender_helo_name

 deny
   condition = ${if eq{$sender_helo_name}{$interface_address}{yes}{no}}
   message = You are not $sender_helo_name I am $interface_address
   log_message = Tu no eres $sender_helo_name

 deny
   !hosts = +relay_from_hosts
   condition = ${if match_domain{$sender_helo_name}{+local_domains}{yes}{no}}
   message = You are not $sender_helo_name I am
   log_message = $sender_helo_name es una IP local

deny
   !hosts = +relay_from_hosts
   condition = ${if isip{$sender_helo_name}{yes}{no}}
   condition = ${if eq{$sender_helo_name}{$sender_host_address}{no}{yes}}
   message = You are not $sender_helo_name
   log_message = Tu IP no es la misma que la que dices $sender_helo_name

deny
   condition = ${if match {$sender_helo_name}{\N^[A-Z0-9]+\.[a-z]+$\N}{yes}{no}}
   condition = ${if match {$sender_helo_name}{\N^[0-9]+\.[a-z]+$\N}{no}{yes}}
   message = You are not $sender_helo_name
   log_message = Virus HELO grepword $sender_helo_name

deny
   !hosts = +relay_from_hosts
   condition = ${if match {$sender_helo_name}
{(backup.lst|localhost.localdomain)}{yes}{no}}
   message = $sender_helo_name does not exist
   log_message = HELO no existente $sender_helo_name


deny
   !hosts = +relay_from_hosts
   condition = ${if eq{$sender_helo_name}{}{yes}{no}}
   message = HELO esta vacio

deny
   !hosts = +relay_from_hosts
   condition = ${if match {$sender_helo_name}{\N^\[.*\]$\N}{yes}{no}}
   condition = ${if eq{$sender_helo_name}{[$sender_host_address]}{no}{yes}}
   message = You are not $sender_helo_name
   log_message = [IP]!= real IP
```

Después de este punto, se han descartado la mayoría de los correos generados

por *spammers*. Corresponde ahora afinar un poco más las reglas evitando por ejemplo caracteres no permitidos.

```
deny
  local_parts = ^.*[@%!/|] : ^\\.
  message = Denegado por localparts

deny
  domains = +local_domains
  local_parts = CHECK_RCPT_LOCAL_LOCALPARTS
  message = restricted characters in address

deny
  domains = !+local_domains
  local_parts = CHECK_RCPT_REMOTE_LOCALPARTS
  message = restricted characters in address


warn
  message = X-Host-Lookup-Failed: RevDNSlookup failed for $sender_host_address
            (${if eq{$host_lookup_failed}{1}{failed}{deferred}})
   condition = ${if and{{def:sender_host_address}{!def:sender_host_name}}\
                {yes}{no}}
```

Se chequea que el *server* que está intentando enviar no esté en alguna lista negra o reportado como *open relay*.

```
drop
  !hosts        = +relay_from_hosts
  message       = found in $dnslist_domain
  dnslists      =  in.dnsbl.org/$sender_address_domain : \
                  list.dsbl.org : \
                  relays.ordb.org : \
                  sbl-xbl.spamhaus.org
```

Se aceptan los correos provenientes de los *servers* a los que se le da servicio de salida.

```
accept
  hosts = +relay_from_hosts
```

Se hace una verificación de la persona que esta enviando el correo.

```
accept
  domains = +local_domains
  verify   = sender/callout
```

Se niega el resto.

```
deny
  message = relay not permitted
```

Después de estas reglas la mayoría de los correos que no son válidos quedan descartados. Toca entonces el turno al *spamassassin* de validar si el correo es un SPAM o no.

Para eso se acepta el cuerpo del correo y se le pasa al programa spamc que es el que devolverá una clasificación de SPAM o no. De esa clasificación dependerá que el EXIM deje seguir al correo o simplemente lo borre.

3.4 Análisis económico

El análisis económico del las pérdidas que representaba el SPAM para la UCLV queda descrito en la tabla 3.1.

Las cifras demuestran que más que una necesidad, era una urgencia la realización de alguna medida para combatir el fenómeno, surgiendo así la idea de la realización de este proyecto que monetariamente no requería de inversión alguna siendo completamente rentable, pues solo se utilizan las potencialidades de *software* libres, como el Exim y el *spamassassin* para encontrar la solución óptima y así disminuir las pérdidas a la institución por concepto de SPAM.

3.5 Al final de todo

El Exm4 es un MTA estable y flexible creado para sistemas UNIX. Permite poner reglas con buen resultado en el filtrado SMTP y así como la integración con *spamassassin,* que a la vez permite una configuración para varios tipos de reglas creando una herramienta fortísima contra el SPAM.

La herramienta que se propone para atenuar el problema del SPAM existente en la UCLV es la integración del Exim4 y el *spamassassin*.

Tabla 3.1. Pérdidas por concepto del SPAM en la UCLV

Coste del SPAM				
Número de empleados	1062	empleados	Coste por hora trabajada	2.65$
Correos diarios recibidos por la organización	21240	correos al día	Coste por minuto trabajado	0,04417$
Salario medio anual de un empleado	6370,8$		Coste por día trabajado	21,24,$
Correo diario por empleado	20		Número de días laborables	261
Porcentaje de SPAM recibido	30%	(estimado)	Horas a la semana	40
SPAM recibido	6	correos	Horas trabajadas al día	8
Tiempo para leer y borrar un correo SPAM	10	segundos	Horas trabajadas al año	2088
Pérdida de productividad diaria por empleado	0,01667	horas		
Pérdidas en productividad anual por empleado	4,35	horas		
Días de productividad perdidos en la empresa	192,49	días		
Coste de pérdidas por empleado anual	11,52$			
Pérdidas de productividad anual	**12242,21$**			
Tasa de respuestas	1,00%		Ancho de Banda (BW)	1 MB
Correos SPAM respondidos	0,06	correos	Precio mensual	1300 c.u.c
Tiempo para responder un correo	3	minutos	BW ocupado por SPAM en %	30%
Pérdida de productividad diaria por empleado	0,003	horas	Pérdidas mesual	390 c.u.c
Pérdidas en productividad anual por empleado	0,783	horas	Perdidas anules	**4680 c.u.c**
Días de productividad perdidos en la empresa	34,65	días		
Coste de pérdidas por empleado anual	2,07$			
Coste de la respuesta al SPAM anual	**2203,6$**			
Coste total para la organización anual (MN)	**14445,81$**		Coste total para la organización anual (c.u.c)	**4680 c.u.c**

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

Tras el cumplimiento del objetivo principal de este trabajo diploma y mediante la realización de las tareas técnicas se concluye que:

1 El SPAM es un verdarero problema para el funcionamiento óptimo y rentable de las redes conectadas a Internet.

2 La red de la UCLV no escapa a la problemática mundial asociada al SPAM.

3 Para combatir el problema se debe garantizar la correcta selección, implementación y configuración de la herramienta para el filtrado, así como la preparación del personal asociado a esta tarea.

4 No es posible eliminar radicalmente el SPAM, pero sí se pueden tomar varias medidas para atenuar su incidencia.

Recomendaciones

1 Actualizar constantemente las reglas del Exim4 y del *spamassassin* debido a que las mismas pueden caducar en pocos días.

2 Profundizar en el estudio de las reglas Exim4 para que el solo sea capaz de detectar qué correo es SPAM, sin la ayuda del *spamassassin* mejorando así la disponibilidad del canal y comunicaciones para otras tareas.

3 Identificar a los usuarios de la red con el problema del SPAM y sus concecuencias, así como del correcto uso del correo electrónico.

4 Crear una vía para que los usuarios de la UCLV retroalimenten a los administradores con los correso de SPAM que los reciben de forma que se puedan ajustar mejor los clasificadores.

REFERENCIAS BIBLIOGRÁFICAS

Cea, J. (2007) Mis experiencias con SPF (Sender Policy Framework), http://www.argo.es/~jcea/antispam/spf.htm

Chacón, L. (2006) *Reforma.com*, http://www.ironport.com/ar/company/reforma_29-11-06.html.

Chávez, C. (2000) ¿Qué es correo electrónico?, http://sisbib.unmsm.edu.pe/bibVirtual/Publicaciones/quipukamayoc/2000/segundo/correoelectronico.htm

Committee, P. M. (2007) The Apache spamassassin Project, http://spamassassin.apache.org/

Cortés, A. (2002) *noticiasdot.com*, http://www.noticiasdot.com/publicaciones/2002/1102/081102/noticias081102/noticias081102-19.htm.

De CocK, J., Bhagat, D. yWright, T. (2006) http://www.tldp.org/HOWTO/Spam-Filtering-for-MX.

Fernández, J. (2005) *Germinus*, http://www.germinus.com.

García, A. (2001) http://www.geocities.com/bertogg/linux/exim-como.html.

Graphics, S. L. (2007) Enciclopedia Gráfica, http://www.sitographics.com/enciclog/banderas/entrada.html

Harris, E. (2004) Greylisting, http://projects.puremagic.com/greylisting/index.html

Hauben, M. (2004) History of ARPANET, http://www.dei.isep.ipp.pt/~acc/docs/arpa-Contents.html

IETF, G. d. t. d. l. (2004) MTA Authorization Records in DNS (marid), http://www.ietf.org/html.charters/OLD/marid-charter.html

Kalinin, A. yVlosava, A. (2007) *Kaspersky Security Bulletin*, http://viruslist.com.

López, Y. (2007), Documentación de la Red UCLV.

Lundgren, B. (2004) Greylisting, http://www.greylisting.org/

MailxMail.com (2007) SPAM, http://www.mailxmail.com/curso/informatica/spam

Marcelo.ar (2005) Microsoft demanda a distribuidores de correo basura mediante "Spam

Zombies", http://marcelo.zoomblog.com/archivo/2005/10/

Mehnle, J. (2006) SPF FAQ (preguntas frecuentes sobre SPF), http://spf.pobox.com/faq.html

Menéndez, C. (2004) Historia del correo electrónico, http://www.telecable.es/personales/carlosmg1/historia_correo.htm

Microsoft, C. (2007) Sender-ID, http://www.microsoft.com/senderid

Naumov, V. (2006) *viruslist.com*, http://www.viruslist.com/sp/spam/analysis?pubid=207270893.

Osmosis, L. (2007) Servidor-Mail, http://www.osmosislatina.com/soporte/servidor_mail.htm

Padrón, C. (2007) 12 de abril de 1994: Primeros correos SPAM de Internet, http://padronelpaso.net/b2evolution/index.php?title=12_de_abril_de_1994_primeros_correos_spa&more=1&c=1&tb=1&pb=1

Panda, S. (2007) Spam: mensajes de correo no solicitados, http://www.pandasoftware.es/virus_info/spam/

Postel, J. (1975) RFC 706 On the Junk Mail Problem, http://www.rfc-archive.org/getrfc.php?rfc=706

Project, T. S. (2007) http://www.spamhaus.org.

Querétaro, I. T. d. (2006) El servidor DNS, http://www.itq.edu.mx/vidatec/espacio/aisc/windowsnt/ServidorDNS.html

Smaldone, J. (2006) http://blog.smaldone.com.ar/2006/12/05/como-funciona-el-dns.

Society, I. (2001) Search the RFC Index, http://www.rfc-editor.org/rfcsearch.html

Spamhaus, T. P. (2004) Register of Known Spam Operations (ROKSO), http://www.spamhaus.org/rokso/index.lasso

Technology, C. f. D. a. (2003) Why Am I Getting All This Spam:Unsolicited Commercial E-mail Research Six Month Report, http://www.cdt.org/speech/spam/030319spamreport.shtml

Templeton, B. (2004a) Origin of the term "spam" to mean net abuse, http://www.templetons.com/brad/spamterm.html

Templeton, B. (2004b) Reaction to the DEC Spam of 1978, http://www.templetons.com/brad/spamreact.html

Wikipedia, E. l. (2007a) Entrada de Canter y Siegel, http://en.wikipedia.org/wiki/Canter_&_Siegel

Wikipedia, E. l. (2007b) Entrada de DNSBL, http://en.wikipedia.org/wiki/DNSBL

Wikipedia, E. l. (2007c) Entrada del término SPAM asociado al correo electrónico, http://en.wikipedia.org/wiki/E-mail_spam

Zenger, R. (2004) What do you get when you buy a spam CD, http://rejo.zenger.nl/abuse/emailcd.php

ANEXOS