

¿Control de contenido? Bien, gracias

El crecimiento de Internet y la inabarcable cantidad de información disponible en la Red hacen de la navegación una verdadera aventura con desenlace incierto. Lo único cierto es que una buena parte del ámbito empresario toma nota de las potenciales consecuencias derivadas de la conducta de sus empleados en el mundo Web. Algo similar ocurre en organismos públicos y oficinas estatales. Si bien es importante contar con un marco normativo basado en Políticas de Seguridad Informática que definan, entre otras cosas, parámetros de uso aceptable de los servicios, el acceso masivo a servicios de Internet se puede circunscribir de manera efectiva mediante herramientas de control de contenido, con resultados y beneficios tangibles. A nivel hogareño, el principal desafío es garantizar un acceso “sano” a los menores de edad.

Telaraña peligrosa

Desde hace algún tiempo, muchas entidades especializadas advierten sobre la cantidad y calidad de información que circula a diario en diversos formatos, ya sea mediante correo electrónico o a través de sitios Web. Un flagelo como el spam se encarga incesantemente de transportar desde inocuas ofertas sobre medicamentos mágicos, pasando por campañas masivas que incluyen pornografía, pedofilia, hasta terrorismo. Si a esto se suman estudios que le ponen precio a las horas de trabajo perdidas y calculan el dinero esfumado cada vez que alguien hace clic en un browser, el panorama es bastante complejo, por no decir caótico.

En un mundo cada vez mas conectado, impedir de plano que los integrantes de una organización puedan “surfear en el ciberespacio” no aparece como una solución original, ya que implica aislarlas y evitar su interacción con servicios como portales, noticias, home banking, etc., por citar algunos. En el otro extremo aparece el acceso indiscriminado al ingente volumen de material que ofrece el universo virtual. El punto medio entre ambas situaciones supone la búsqueda de un equilibrio difícil de lograr y que debe seguir un rumbo bastante tortuoso: definir estrategias concretas, conseguir apoyo de los niveles directivos, generar Políticas y Normas formales, para recién luego implementar soluciones informáticas que soporten la estructura previamente diagramada.

El contenido bajo la lupa

Una de tales soluciones que permiten proteger al internauta y mejorar su experiencia en cuanto al uso de la Web y el correo electrónico es el control o filtrado de contenido, del cual intentaré explicar brevemente sus principios básicos. Antes de entrar al detalle, debe quedar en claro que este tipo de implementaciones requiere una combinación de hardware, software y bastante trabajo de configuración, además de un monitoreo constante de las prestaciones.

El filtrado de contenido cubre dos áreas bien delimitadas: negocios e Internet. También pueden identificarse tres destinatarios básicos para instalar este tipo de herramientas: empresas, proveedores de acceso y usuarios hogareños. En entornos corporativos los mejores resultados se obtienen mediante un esquema cliente – servidor. Del lado del servidor los filtros de contenido pueden actuar de dos maneras:

- a) Mediante una aplicación de software instalada en un equipo (PC o servidor) dedicado que actúa como proxy. En este caso la aplicación se encarga de filtrar todo el tráfico de la red que circula a través del equipo host. En algunas situaciones el software admite la configuración de dos interfaces, lo cual permite filtrar y monitorear el tráfico que pasa de una placa a la otra.
- b) Usando un dispositivo de hardware conocido como *network attached appliance*, que también actúa como gateway y proxy, pero que es bastante más costoso económicamente y posee un rendimiento superior comparado con la opción a).

¿Cómo funciona? Listas negras y heurística

La mayoría de las soluciones de filtrado de contenido trabajan sobre tráfico HTTP. El puerto por defecto usado por este protocolo es el 80, aunque es posible configurar puertos alternativos tales como el 8080 y el 3128, dependiendo de las necesidades específicas.

Desafortunadamente, el tráfico HTTP no es la única manera de recibir datos en una PC. Aun si se utilizan túneles encriptados, existen muchos otros servicios que deben tenerse en cuenta, a saber: intercambio de archivos mediante P2P, recepción de adjuntos en correo electrónico, mensajería instantánea, descargas usando protocolos como NNTP o FTP, uso de programas tipo Skype; etc. La enumeración suele ser más larga de lo que un administrador pretende.

Las listas negras son, esencialmente, bases de datos que contienen nombres de dominios y direcciones de correo electrónico que por diversos motivos (servidores mal configurados, spam, open relay) no son fiables. Existen empresas dedicadas a crear y mantener black lists, cuyo tamaño en la actualidad supera los 20 Gb., aunque parezca mentira. En el caso del software de filtrado, tales bases de datos vienen preinstaladas y se actualizan diariamente. También es posible agregar registros en forma manual y tener una lista negra personalizada, lo cual supone una tarea administrativa tediosa pero necesaria, sobre todo al principio de una implementación. La contrapartida está dada por las listas blancas, que permiten asegurar el acceso a determinados sitios y garantizan, entre otras cosas, la llegada de correos de ciertos dominios.

La heurística es un complemento para el bloqueo de contenido. Se trata de una tecnología que procesa conjuntos de reglas aplicadas al tráfico de red. El uso más difundido de reglas es la detección de palabras, tanto en la navegación como en el cuerpo de mensajes de correo electrónico. Las reglas admiten la asignación de un “peso” o “sensibilidad” a cada término evaluado; de esta manera es posible establecer valores en escalas de 1 a 5, o de 1 a 100, dependiendo del grado de flexibilidad que ofrezca la herramienta. La principal crítica que se le puede hacer a esta metodología es que aún no es posible analizar cada palabra dentro del contexto en el que se utiliza; a fin de evitar el bloqueo de texto inofensivo. No es lo mismo hacer una búsqueda de la frase “oferta sexual” que “educación sexual” o “enfermedades de transmisión sexual”. No obstante, la heurística ha tenido algunos avances a la hora de controlar patrones, examinar texto incorporando diccionarios en varios idiomas y diferenciar cadenas alteradas, por ej. “hardcore”, “hard_core”, “hard0core”, “HARDc0r” y las combinaciones posibles de esta palabra. Algunos de los productos mas difundidos son CA eTrust, Content Keeper, Internet Sheriff, Surf Control, Websense, Cyber Patrol, Net Nanny, Pure Sight, Dans Guardian.

Y por casa, ¿cómo andamos?

Pensando en el uso hogareño, y dado que no siempre es posible disponer de software específico, los navegadores de Internet más conocidos incorporan asesores de contenido propios. Si bien son bastante básicos y rudimentarios, permiten hacer ciertas operaciones como clasificar la información por categorías (desnudez, lenguaje, sexo, violencia, alcohol, tabaco, armas, etc.), armar una lista de sitios aprobados y no permitidos, definir una contraseña de supervisión y adoptar algún sistema de clasificación mundialmente aceptado y reconocido, por ej. las iniciativas RSACi (www.rsac.org), PICS (www.w3.org/pics) o ICRA (www.icra.org/pics) que ayudan a proteger a los menores.

Conducta vs. Tecnología

Al igual que en otras disciplinas, no debe perderse de vista que tanto Internet como la infraestructura que la soporta configuran un abanico de servicios y recursos que no puede regularse ni gobernarse en forma rígida. De alguna manera la Red se ha convertido en una maraña acéfala donde la tendencia marca el camino hacia la anarquía, a pesar de los proyectos y esfuerzos por lograr una administración consensuada, con reglas claras. Mas allá de las áreas que puedan ser cubiertas por el control de contenido (mensajería instantánea, Webmail, home banking, apuestas deportivas, juegos, búsquedas laborales, publicidad, etc.) no debe obviarse el concepto de “identidad digital”. Muchas personas piensan que el “anonimato” del cual creen gozar les permite usar la Web y el correo electrónico para hacer cosas que no se conciben con su conducta en la vida real. Algunas empresas, a través de políticas escritas que se entregan a los empleados, equiparan la actuación de los mismos en la Web a la presencia personal (en vivo y en directo), considerando que el individuo en relación de dependencia representa a la organización mientras navega o cada vez que envía un mensaje instantáneo. Es importante situar el sentido común y una conducta responsable y coherente antes que la tecnología. Después de todo, el mundo virtual no está tan lejos del mundo real.

Walter Heffel
Noviembre de 2006