

Este artículo pretende poner de manifiesto que los motores de búsqueda pueden utilizarse como excelentes herramientas de 'hacking'. Este hecho se debe, sobre todo, a la acusada falta de adecuación de las políticas de seguridad que implementan. Por esta razón es relativamente sencillo abusar de los servicios que ofrecen, por ejemplo para atacar máquinas anónimamente, buscar víctimas vulnerables, u obtener datos confidenciales.

## Buscadores y seguridad

El objetivo principal de este artículo es ilustrar la importancia que para todo usuario de Internet tendría el incremento del nivel de seguridad de estos motores de búsqueda. Es evidente que hay un gran número de máquinas con una seguridad muy débil en Internet pero, lo que es aún más grave es que existen métodos anónimos muy sencillos para detectarlas y abusar de ellas. El texto siguiente se centra en algunos de estos métodos, y específicamente en aquellos basados en los motores de búsqueda. Además, se mencionan algunas de las medidas que pueden implementar tanto los administradores de equipos conectados a la red como los de los buscadores, para aumentar su seguridad y, por tanto, dificultar las actividades de posibles atacantes.

### INTRODUCCIÓN

Los motores de búsqueda en Internet ofrecen, cada vez más, un mayor número de servicios, e indexan un gran número de páginas web y otros recursos que algunas veces, inadvertidamente, exponen las debilidades de seguridad de las compañías, e incluso sus datos confidenciales.

En este trabajo se presentan algunas técnicas que demuestran que los hackers pueden abusar de los buscadores para realizar anónimamente sus ataques, encontrar víctimas fáciles y adquirir información confidencial que, en algunos casos, pueden ser más que suficientes para llevar a término un potente y prometedor ataque.

El anonimato es el primer objetivo de cualquier atacante, tanto para eludir las consecuencias legales de sus actos, como para, simplemente, evitar ser rechazados por sus Proveedores de Servicios de Internet. Puede abusarse de algunos servicios que los motores de búsqueda ofrecen para forzar a que actúen como proxies anónimos, consiguiendo de esta manera ocultar la identidad de los usuarios con intenciones poco confesables.

Hay muchas personas que no son realmente conscientes de las implicaciones que tiene la existencia de máquinas con una seguridad muy débil en Internet. Estas máquinas pueden ser utilizadas para comprometer la seguridad de cualquier otra máquina conectada a Internet de varias maneras.

Algunas veces estos servidores no almacenan información importante o confidencial, pero mantienen una relación de confianza con terceras redes. Consecuentemente, un atacante puede acceder con facilidad a una de estas máquinas vulnerables y, desde ésta, atacar una importante red confiable. De esta manera, el atacante puede obtener, además, el anonimato una vez haya destruido los ficheros de log o cualquier pista en la máquina vulnerada. Otra manera de explotar estas máquinas es lanzar ataques distribuidos de Denegación de Servicio (DDoS), transformándolas en máquinas zombies. Afortunadamente, existen una serie de contramedidas para mitigar estas amenazas que, si fueran implementadas por los motores de búsqueda, dificultarían en gran medida la posibilidad de abusar de las máquinas vulnerables.



Figura 1: Comprobación del anonimato obtenido por el motor traductor de Altavista. La dirección IP mostrada en el campo REMOTE\_ADDR pertenece a una máquina de Altavista, así como el resto de los campos, permitiendo, por tanto, que un usuario abuse de este servicio para navegar anónimamente.

Algunas veces estos servidores no almacenan información importante o confidencial, pero mantienen una relación de confianza con terceras redes. Consecuentemente, un atacante puede acceder con facilidad a una de estas máquinas vulnerables y, desde ésta, atacar una importante red confiable. De esta manera, el atacante puede obtener, además, el anonimato una vez haya destruido los ficheros de log o cualquier pista en la máquina vulnerada. Otra manera de explotar estas máquinas es lanzar ataques distribuidos de Denegación de Servicio (DDoS), transformándolas en máquinas zombies. Afortunadamente, existen una serie de contramedidas para mitigar estas amenazas que, si fueran implementadas por los motores de búsqueda, dificultarían en gran medida la posibilidad de abusar de las máquinas vulnerables.

### USO DE LOS BUSCADORES COMO PROXIES ANÓNIMOS

Multitud de motores de búsqueda, incluyendo dos de los más populares, Altavista y Hotbot, ofrecen el servicio de traducción automática de páginas web a sus usuarios. La forma de trabajar de este servicio es simple: el usuario solicita una URL a los motores de búsqueda con traducción automática, éstos descargan el recurso localmente, lo traducen, y finalmente devuelven el resultado al usuario.

Así pues, este procedimiento permite a cualquier usuario utilizar eficazmente el motor traductor como proxy; esto puede considerarse un hecho inofensivo si no fuera porque tanto Altavista como HotBot actúan, además, como proxies anónimos. De esta forma se permite que cualquier usuario



Figura 2: Los servicios de traducción ofrecidos por Altavista, HotBot y otros se pueden utilizar no sólo para encontrar máquinas vulnerables (por ejemplo, aquellas que poseen vulnerabilidades cgi), sino también para explotarlas anónimamente.

pueda navegar anónimamente, ocultando su dirección IP auténtica detrás de la dirección IP del motor traductor (Figura 1).

También es posible y extremadamente fácil encadenar varios buscadores con traductores. De esta manera se provoca que la identificación del usuario sea aún más difícil.

Aunque la navegación de forma anónima puede no parecer un riesgo serio, este anonimato se puede utilizar para evitar los efectos legales que algunas simples, pero malévolas, peticiones http pudieran tener. Por ejemplo, la petición http

`http://www.some.web.server/msadc/Samples/SELECTOR/showcode.asp?source=/msadc/Samples/./././././boot.ini` devolvería los contenidos del fichero boot.ini en algunos servidores web que ejecuten Internet Information Server (IIS) 3.0 y 4.0 (Figura 2); por tanto, esta característica de los principales motores de búsqueda puede utilizarse para encontrar y explotar anónimamente vulnerabilidades de los servidores web.

### BÚSQUEDA DE VÍCTIMAS FÁCILES

La enorme cantidad de páginas indexadas por los mayores buscadores (por ejemplo, Google anunciaba 1.346.966.000 en julio de 2001) convierten a éstos en excelentes herramientas para identificar víctimas fáciles.

Una técnica muy simple y extendida es localizar servidores web recién instalados o desatendidos, ya que con gran probabilidad no poseen una buena seguridad. Esto se puede llevar a cabo sencillamente buscando frases, imágenes u otros contenidos que caractericen la instalación predeterminada de ciertos servidores web. Por ejemplo, si la



Figura 3: Página web predeterminada de una instalación de IIS 3.0 con la frase propuesta resaltada. Esta frase propuesta se puede ampliar o modificar para caracterizar mejor las instalaciones básicas de servidores web.

frase “Try the hyperlinks above to see some examples of the content you can publish with Microsoft Internet Information Server. To learn more about Microsoft products that you can use to create great-looking Web pages” se encuentra en un servidor web, hay un alta probabilidad de que se trate de una instalación predeterminada de IIS (Figura 3).

Análogamente, la frase “This page is used to test the proper operation of the Apache Web server after it has been installed. If you can read this page, it means that the Apache Web server installed at this site is working properly.” caracteriza una instalación básica de un servidor Apache (Figura 4).

Este método para encontrar instalaciones desatendidas o recientes proporciona al hacker una larga lista (Tabla 1) de potenciales víctimas fáciles de atacar.

Los intrusos también pueden intentar localizar páginas por defecto correspondientes a versiones antiguas de servidores web (con vulnerabilidades bien conocidas), simplemente incluyendo la fecha de copyright en la frase buscada (como en “©1997 Microsoft Corporation. All rights reserved”). Otra posibilidad es buscar servidores web que tengan una reputación muy deficiente en cuanto a seguridad se refiere, como es el caso del Personal Web Server de Micro-



Figura 4: La frase propuesta para la instalación básica de un servidor Apache aparece resaltada.

soft (que podría buscarse utilizando la frase “This home page is hosted on Microsoft Personal Web Server (PWS) 4.0. PWS turns any computer running Windows® 95 or Windows 98 into a Web server and enables instant publication of personal Web pages”), o el del servidor web Zeus.

Obviamente, el método de búsqueda expuesto en los párrafos anteriores puede ser ligeramente mejorado ampliando la frase a buscar o restringiendo la búsqueda sólo a páginas web recientemente indexadas. Se pueden aplicar otras modificaciones, pero la discusión de cómo obtener mejores resultados (o víctimas más indefensas) está fuera de los objetivos de este artículo.

	Instalaciones predeterminadas de Apache indexadas	Instalaciones predeterminadas de IIS indexadas
www.google.com	9360	2970
www.lycos.com	2310	105
www.altavista.com	6205	3824

Tabla 1: Número de respuestas obtenidas tras una petición de búsqueda con las frases propuestas para IIS y Apache.

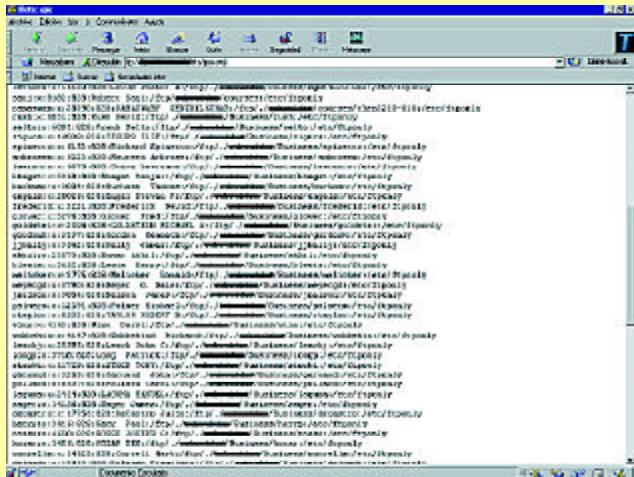


Figura 5: Ejemplo de fichero /etc/passwd obtenido mediante una petición al motor de búsqueda de ficheros de Lycos. Esta información se podría utilizar para realizar un ataque de ingeniería social.

Alternativamente, muchos de los motores de búsqueda ofrecen a sus usuarios la localización de contenidos multimedia tales como gráficos, banners o logotipos, incluyendo ficheros de imágenes (jpeg, gif, etc.) presentes en las páginas web de las instalaciones predeterminadas. Este tipo de búsquedas, por sí mismas o en combinación con los tipos de búsqueda de frases textuales anteriormente descritos, resultan especialmente útiles para localizar instalaciones de servidores web vulnerables.

### BÚSQUEDA DE DATOS Y ARCHIVOS CONFIDENCIALES

Se ha expuesto, esperamos que claramente, cómo algunos motores de búsqueda son una amenaza seria de seguridad para el resto de máquinas conectadas a Internet debido a sus políticas de indexación y a la falta de defensas contra usuarios maliciosos. Si los motores de búsqueda convencionales adolecen de los citados problemas, el caso de los motores de búsqueda de ficheros es aún más peligroso.

Los motores de búsqueda de ficheros, como es el caso del buscador de Lycos, tienen miles de entradas relativas a máquinas débiles, máquinas distribuidas por todo Internet que filtran información confidencial.

Incluso si se utiliza el cifrado de la información, ningún administrador sensato desearía que los ficheros de contraseñas de sus máquinas pudieran indexarse desde un motor de búsqueda, provocando que todo el mundo tuviera acceso a ellos. Asombrosamente, eso es exactamente lo que ocurre en miles de máquinas de la red.

Obviamente, los motores de búsqueda de ficheros no son los únicos responsables de esta fuga de información, que se podría evitar si las máquinas se protegieran adecuadamente.

Seguidamente se citan algunos ejemplos:

Usando Lycos como motor para buscar ficheros tan confidenciales como pueden ser /etc/passwd y /etc/shadow, se obtienen cientos de interesantes resultados como los que se muestran en las figuras 5 y 6.

No sólo ficheros tan comunes como /etc/passwd y /etc/shadow son fuente de información importante acerca de las máquinas. Por ejemplo, los ficheros .htaccess y .htpasswd, usados respectivamente para controlar el acceso a ciertos contenidos de un servidor web y para almacenar las contra-

señas que autorizan estos accesos, son también relevantes y muy fáciles de encontrar con la inestimable ayuda del buscador de Lycos. En las figuras 7 y 8 se muestran dos ejemplos.

En muchos casos, la explotación de la información que filtran los motores de búsqueda es tan simple y rápida como en el caso en el que un atacante encuentra ficheros cifrados con un algoritmo poco robusto. Este escenario le permite recuperar una contraseña casi instantáneamente. Las contraseñas se reutilizan con frecuencia, y esto puede facilitar en gran medida la tarea de adivinar la contraseña de otros usuarios, incluso root. Un ejemplo típico de algoritmo poco robusto podría ser el de cifrado que utiliza CuteFTP para almacenar las contraseñas en el fichero SMDATA.DAT, o el de Netscape Enterprise Server que guarda las contraseñas en el fichero admpw. Una búsqueda de ficheros del tipo SMDATA.DAT o admpw permitirá a un atacante obtener información muy sensible de manera muy simple y rápida.

### BÚSQUEDA DE RESULTADOS DE AUDITORÍAS

Es importante resaltar, para terminar esta exposición de los riesgos de los motores de búsqueda, que éstos también pueden ser utilizados para encontrar resultados de auditorías y valoraciones sobre las máquinas, datos que proporcionan información crucial a un atacante potencial.

Un interesante ejemplo de lo anterior es el siguiente fichero que se puede obtener tras realizar una sencilla búsqueda en Google (Figura 9). Este fichero contiene auditorías de más de 4900 servidores web, todos ellos en el dominio \*.edu, mucho de los cuales tienen una seguridad pésima.

### ALGUNAS POSIBLES CONTRAMEDIDAS

Claramente, los motores de búsqueda deberían tomar medidas para evitar que pudieran indexar información confidencial. Dejando al margen el hecho de que ayuda a los posibles atacantes, este comportamiento podría ser considerado como acto criminal bajo ciertas legislaciones. De todas maneras, actualmente los buscadores no realizan ninguna acción para contrarrestar estas filtraciones indeseables de

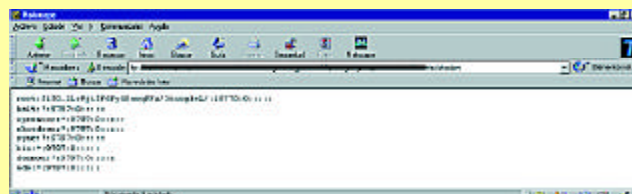


Figura 6: Ejemplo de fichero /etc/shadow obtenido mediante una petición al motor de búsqueda de ficheros de Lycos. Este fichero cifrado, con la contraseña de root, se puede utilizar para realizar con éxito un rápido ataque de diccionario.

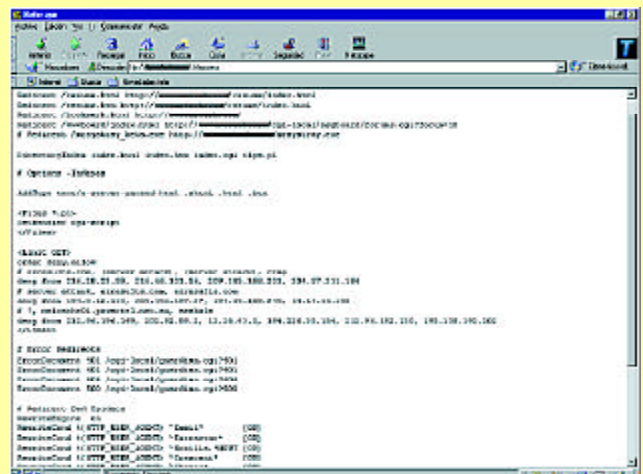


Figura 7: Ejemplo de fichero .htaccess obtenido mediante el motor de búsqueda de ficheros de Lycos. El fichero contiene información acerca de redirecciones y control de accesos que rechazan ciertas direcciones.

información, lo que no puede considerarse tan sólo culpa de la política de los webbots sino también de los administradores de las máquinas.

Afortunadamente, pueden aplicarse unas cuantas ideas, fáciles de llevar a cabo, que resuelven, o al menos minimizan, estas filtraciones de información.

A continuación se presentan algunas de ellas:

1) La solución en la que los traductores se comportan como proxies es la más natural y simple si no fuera por el hecho de que pueden actuar como proxies anónimos. Esta característica de anonimato sería un problema que se solucionaría fácilmente si el traductor comunicara al servidor cuál es el origen real de la petición. Esto se puede llevar a cabo, por ejemplo, utilizando los campos REMOTE\_ADDR, HTTP\_X\_FORWARDED\_FOR o HTTP\_VIA. Si esto se implementase en todos los motores de búsqueda, los intrusos potenciales dejarían inmediatamente de utilizarlos como proxies anónimos. Si, por el contrario, esto no se implementa nunca, probablemente aumentaría el número de personas que utilizaran este mecanismo, y más administradores de red prohibirían el acceso a sus páginas desde direcciones IP que contuvieran traductores. El caso de Infoseek (también conocido como Go) es un ejemplo inusual de buena

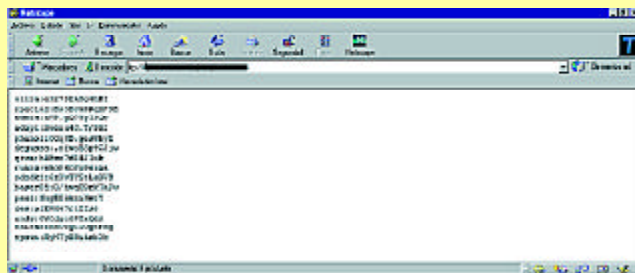


Figura 8: Ejemplo de fichero .htpasswd obtenido mediante el motor de búsqueda de ficheros de Lycos. Esta información se puede utilizar para realizar un ataque de diccionario mediante Crack o John the Ripper tras ligeros cambios de formato.

práctica: aunque proporcionan servicios de traducción a través de SystranSoft, no es posible abusar de ellos porque utilizan el campo http REMOTE\_ADDR para informar al servidor web del origen real de la petición http. También evitan el encadenamiento de traductores que se comportan como proxies rechazando peticiones de auto-traducción.

2) No hay ninguna razón para indexar las páginas que ofrecen los servidores web en sus instalaciones básicas. Esta indexación sólo es útil para encontrar víctimas fáciles, así que deberían programarse los webbots para no devolver páginas web de instalaciones básicas o a los motores de búsqueda para no mostrar este tipo de resultados. Ambos, webbots y motores de búsqueda lo tendrían más fácil si los desarrolladores de servidores web incluyeran un archivo restrictivo por defecto robots.txt, o una etiqueta especial META, por ejemplo:

```
<meta name="function" content="Default Web Server Page">
```

3) Hay multitud de archivos de diferente valor en cuanto a la seguridad que no deberían ser explorados por ningún webbot. De hecho, estos bots deberían seguir una simple política de seguridad: recuperar sólo lo que se encuentre en el directorio /pub de los servidores de ftp, ni más, ni menos. Recuperar, incluso parcialmente, los contenidos del directorio /etc puede ser motivo de persecución legal bajo ciertas legislaciones. El actuar de acuerdo con un fichero similar a /robots.txt puede ser muy útil en el caso de bots de ftp, si los desarrolladores de demonios de ftp incluyeran uno en sus distribuciones. Desafortunadamente, éste no es el caso y llevará algún tiempo (y algún que otro incidente de seguridad) llegar hasta ahí.

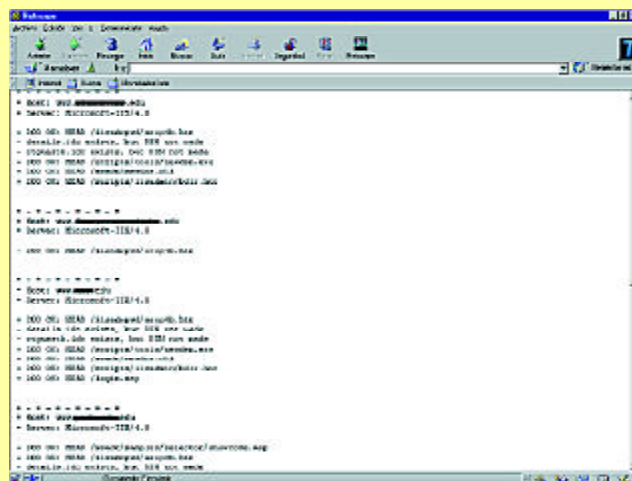


Figura 9: Parte de un fichero mostrando miles de servidores web vulnerables en el dominio \*.edu.

## CONCLUSIONES

Como se ha expuesto, en este artículo se presentan posibles mejoras en la seguridad de Internet. Está claro que no es una tarea trivial y que los problemas descritos en las anteriores secciones son sólo una pequeña muestra de ellos.

Los motores de búsqueda se han convertido en una interesante herramienta de hacking y, como cualquier nueva y potente herramienta de hacking, los administradores de seguridad deberían estar al tanto de sus usos.

Obviamente, toda la responsabilidad no debería recaer sobre los motores de búsqueda: no son culpables del gran aumento de máquinas inseguras en la red. En cualquier caso, se deberían emprender algunas acciones de forma inmediata para imposibilitar, o dificultar, el abuso de estos servicios, que son utilizados para localizar y atacar anónimamente máquinas débiles o información confidencial.

Los autores del presente artículo creen que la presente situación debería mejorarse y recomiendan encarecidamente a los administradores de los motores de búsqueda tener en cuenta las contramedidas propuestas en el mismo. En la actualidad, ningún motor de búsqueda toma medidas para evitar su uso como herramienta de hacking. n

2 Julio César Hernández  
Arturo Ribagorda  
Benjamín Ramos  
Ana Isabel González-Tablas  
Grupo de seguridad  
Departamento de Informática  
UNIVERSIDAD CARLOS III DE MADRID  
{jcesar,arturo,benja1,aigonzal}@inf.uc3m.es

## REFERENCIAS

- Anonymous proxy checker**  
<http://www.multiproxy.org>
- SearchLores**  
<http://www.searchlores.org>
- Seguridad en CGIs**  
<http://secinf.net/info/www/cgi-bugs.htm>
- Algoritmo de cifrado débil en CuteFTP**  
[http://www.securiteam.com/exploits/CuteFTP\\_s\\_password\\_storage\\_insecurity.html](http://www.securiteam.com/exploits/CuteFTP_s_password_storage_insecurity.html)
- Revelación de contraseñas en Netscape Enterprise Server**  
[http://www.securiteam.com/securitynews/Netscape\\_Administration\\_Server\\_Password\\_Disclosure.html](http://www.securiteam.com/securitynews/Netscape_Administration_Server_Password_Disclosure.html)