5. Ingeniería Social

2. Hacking con Buscadores

2.1 Introducción

Cuando la mayoría de los usuarios medios o novatos piensan en internet, la primera imagen que les suele venir a la cabeza es la página principal de Google. Se ha convertido en algo casi cultural, pensamos que Google, lo sabe todo. Y en efecto, hasta cierto punto, es así. Sin embargo tendemos a sobrestimar Google, y subestimar las capacidades de otros buscadores de contenidos como Bing, DuckDuckGo, RTbot, o buscadores de host's como ShodanHQ, de respuestas como Wolfram Alpha, etc.

En este contexto, es importante abrir nuestra mente, y saber que hay alternativas muy útiles, y que en esto del hacking con buscadores, toda ayuda es poca para obtener los resultados que queremos. Si bien un buscador inmenso como Google tiene sus indiscutibles ventajas, las cantidades de información que este maneja a veces pueden ser un hándicap si lo que buscamos es algo muy concreto.

Podemos ver como cada buscador tiene sus propias características, sus puntos fuertes y sus puntos débiles, y conocer una amplia gama de ellos puede ahorrarnos horas de pasar páginas y páginas de Google.

- Google: El gigante de las búsquedas. Google es a día de hoy el buscador con mayor porcentaje de internet en su índice de contenidos. La cantidad de información es ingente, y sin embargo, la búsqueda es rápida. Podemos desde buscar contenidos, hasta buscar por tiempo de aparición, o incluso buscar patrones visuales. Por la gran cantidad de esfuerzos y arañas que tienen indexando internet, muchas veces llegan a los resultados desde contraseñas, hasta bases de datos de usuarios y contraseñas.
- Bing: Bing se caracteriza por acotar los resultados bastante más que Google. Su indexado de cada página es algo más profundo, por lo que habitualmente, indexa más documentos privados que se encuentran en páginas accidentalmente. Otra de las grandes funciones de Bing es la etiqueta "ip:". Con esto podemos saber los nombres de dominio que se resuelven bajo esa misma IP, que en la mayoría de los casos, suelen ser bastantes. Además, Bing tiene una API bastante libre, es por ello por lo que todos los buscadores inversas de DNS suelen basar su funcionamiento en los resultados que ofrece este buscador.
- Shodan: Con el nombre Shodanhq, este buscador ha hecho lo que otros pronto han descartado, indexar host's y sus cabeceras, en vez de contenido. Como resultado tenemos una base de datos a la que hacer consultas con millones de direcciones IP que pueden corresponder a todo tipo de sistemas, desde routers, a sistemas Scada, hasta servidores de backups de empresas, o sistemas utilizados por ISP's para enrutado de paquetes. Además las etiquetas "net:" o la búsqueda por nombres de dominio pueden hacer las delicias de cualquier penetration tester que quiera identificar de manera rápida las IP's utilizadas por, por ejemplo, una universidad.

2.2 Hacking con Google

La principal ventaja de Google, es la cantidad de información que tiene indexada. Podemos decir sin miedo a equivocarnos, que Google es la empresa del mundo con mayor cantidad de información, y como es normal, entre tan inmensa cantidad de datos, hay cosas que no deberían de estar ahí; el Google hacking se basa precisamente en eso, en buscar lo que queremos en medio del caos. Cosa que, con un poco de práctica es más fácil de lo que parece.

Entre toda esa información que no debería estar ahí de la que hablábamos antes, podemos ver cosas como: usuarios y contraseñas en texto plano o hashes, dumps (volcados) enteros de bases de datos de páginas web, logs, mensajes de error, información sensible, código fuente SQL, etc. Toda esta información está ahí por varias razones, desde haber metido una copia de seguridad en la carpeta pública, hasta malas configuraciones del archivo "robots.txt", pasando por bugs, y malas configuraciones.

Es bastante importante saber manejar bien Google (cosa que poca gente sabe hacer), y saber algunos comandos bastante útiles que se usan para este tipo de búsquedas. Por ejemplo: "intitle:","filetype:","ext:","site:","inurl:", además de los clásicos operadores booleanos y caracteres como "+", o "-". Además conviene saber también la "terminología del gremio", es decir cosas como: password, user, root, admin, dump, sql, etc, además de ser aconsejable conocer, al menos a nivel básico, como funciona una web.

Vamos a ver algunos ejemplos:

```
intitle:index.of.config
```

Nos puede mostrar configuraciones de algunos servidores web

```
filetype:bak createobject sa
```

Nos muestra códigos fuente con posibles contraseñas a bases de datos

```
filetype:avastlic
```

Licencias para Avast

```
inurl:"phpmyadmin/index.php" intext:"[ Edit ] [ Create PHP Code ] [ Refresh ]"
```

Bases de datos sin proteger

```
inurl:confidential ext:docx
```

Archivos confidenciales en word (o en cualquier otra extensión)

Y otras búsquedas como usuarios, contraseñas y copias de seguridad de bases de datos de usuarios, enteras:

```
intext:charset_test= email= default_persistent=

ext:sql dump

ext:xml dump

filetype:sql "MySQL dump" (pass|password|passwd|pwd)

filetype:sql inurl:wp-content/backup-*

intext:charset_test= email= default_persistent=
```

Si queremos más dorks, tenemos una buena recopilación aquí: Google-Hacking Database

¿Qué debemos hacer para protegernos de este tipo de cosas?

Si somos administradores: Primero, no subestimar el riesgo pensando: "bah! nadie lo va a encontrar", Google lo hará. Segundo, hashear todas las contraseñas guardadas con un algoritmo decente (lo que excluye MD5), si queremos podemos aplicar varios, o añadir "salt". Tercero, hacer búsquedas de los dorks más comunes sobre nuestra página web. Cuarto, tener el software siempre actualizado. Quinto, tener cuidado en como escribimos código.

Si somos usuarios: Primero, no utilizar la misma contraseña para todo. Segundo, activar el login en 2 pasos de nuestra cuenta gmail (si usamos gmail). Tercero, utilizar contraseñas robustas: mayúsculas, minúsculas, número y símbolos, y siempre de una longitud mayor que 9 caracteres. Cuarto: tener cuidado de donde damos nuestros datos. Quinto: como usuario, podemos exigir legalmente que nuestros datos sean bien custodiados; además, en España hay una agencia de protección de datos que nos escuchará.

2.3 Bing

Si bien Bing no es el buscador en el que pensamos automáticamente cuando hacemos Google-hacking y aunque la cantidad de información bruta sea mucho menor, sí que hay que destacar algunas cosas que este gran buscador nos puede ofrecer, y que otros, no pueden. Los cuatro grandes puntos a destacar sobre Bing cuando hablamos de Hacking con Buscadores son:

- Su búsqueda recursiva de dominios: Como sabemos, puesto que desde hace tiempo hay problemas con la cantidad de ipv4 disponibles, los planes pequeños y medianos de Hosting que ofrecen casi todas las compañías, tienen una IP compartida para un número variable de páginas alojadas. Así, sabiendo una dirección IP, podemos buscar de manera inversa todos los dominios que pueden ser resolubles con dicha IP. Esto se hace de la siguiente manera.

  ```
  ip:82.98.160.207
  ```

- Su profundidad: Si bien es bastante discutible que la profundidad de Bing en cada página web que indexa sea mayor que la de otros buscadores como Google, a la hora de buscar grandes cantidades de documentos dispersos por dominios muy grandes como universidades, u organismos gubernamentales, Bing suele darnos más cantidad de resultados.
- Su extensa API: La API de Bing permite un gran número de consultas diarias, y nos devuelve la información en formato xml. Las consultas pueden ser utilizando cualquier parámetro, es por eso por lo que combinando esto con la búsqueda inversa de DNS y otras herramientas de dominio podemos hacer programas interesantes, análisis de seguridad basados en parámetros específicos del sistema en que están alojadas dichas páginas, etc. Con un poco de creatividad se puede exprimir esta API de formas muy curiosas. Un ejemplo aquí.
- Menos información, pero más relevante: Nadie duda que Google es el gigante de la información, pero esto en muchas ocasiones es un inconveniente. Bing tiene menos información indexada, y para buscar cosas como nombres propios, teléfonos, números de seguridad social, etc, saltarse todas esas páginas llenas de listas interminables de números contiguos es una ventaja.

Sobre los dorks, en Bing casi no hay, porque es un buscador que filtra mucho la información antes de presentarla.

2.4 Shodanhq

Shodanhq es un buscador diferente. Podemos decir que todos los buscadores clásicos parten de la idea de indexar contenido, indexar frases, palabras, ideas, definiciones, imágenes, vídeos, etc. Shodan no busca indexar toda esa información, shodan es básicamente el resultado de pedir las cabeceras a todos los host conectados a internet, a todo el rango de ips: 0.0.0.0/0. Así pues, lo que tenemos es una gran base de datos con, hasta ahora, más de 80 millones de cabeceras de hsots de todo tipo.

Toda esta información vale para algo más que para hacer estadística; podemos hacernos una imagen bastante clara de los sistemas que se utilizan, de sus versiones, pero también de la seguridad de muchos países, organizaciones, y propietarios.

Veamos algunas de las opciones de búsqueda que tiene este buscador.

En este buscador se puede buscar de forma directa, poniendo el nombre de una organización:

```
hostname:google
```

Se puede buscar por rangos de ips:

```
net:208.53.243.0/24
```

Por país (requiere registro):

```
google country:US
```

```
uam country:ES
```

Por puertos:

```
net:209.141.0.0/16 port:22
```

Y muchas otras opciones, aquí tenemos una referencia de los filtros posibles:

http://www.shodanhq.com/help/filters

Y bien... ¿Qué podemos encontrar en Shodan?

Podemos encontrar backups de empresas, ayuntamientos, etc puestos de manera pública y sin ningún tipo de contraseña como ya se explicó en un post anterior; basta con buscar: "iomega".

Podemos encontrar servidores con determinadas versiones con bugs conocidos, como por ejemplo sistemas AirOs de hostings y teleoperadoras donde podemos bypassear el login y obtener acceso root. Podemos encontrar grandes listas de cámaras web emitiendo en abierto por internet. Esto podemos buscar por ejemplo poniendo: Server: SQ-WEBCAM.

Hay también gran cantidad de teléfonos Ips de grandes empresas en los que se podrían interceptar comunicaciones o incluso hacer llamadas gratuitas. Podemos encontrar gran cantidad de routers, la mayoría de los cuales siguen teniendo sus passwords por defecto (otros nos dicen cuál es la contraseña por defecto en la cabecera).

Podemos encontrar sistemas SCADA de control industrial, de control de incineradoras, Ipads, Dreambox, servidores de Spam, sistemas de monitorización de redes, sistemas de control de ISP's, sistemas OsX rooteados con VPN's con contraseñas por defecto, banners curiosos, cámaras londinenses espiando protegiendo a los ciudadanos, gateways de empresas, sistemas de alarmas, bases de datos de empresas sin contraseña, servidores FTP's, semáforos y sistemas de control de semáforos y circulación, sistemas con windows 3.2, routers Wifi, Commodores 64, sistemas de control eléctricos, sistemas de tracking GPS, proxies, lectores RFID de acceso a instalaciones gubernamentales, impresoras, proyectores, servidores de correo electrónico y un montón de cosas que realmente asustan. Aquí tenemos una larga lista.

Si se pasa un rato viendo todos los resultados que se arrojan en cada búsqueda, podemos sacar unas cuantas conclusiones:

- El uso de contraseñas por defecto está extendido, y es uno de los mayores cánceres de la seguridad informática actual.
- Hay mucha información expuesta de manera explícita que en muchos casos es denunciable.
- La política de "qué más da, nadie va a poner justo esa IP", y en general la seguridad por ocultación, no es válida.
- Hay muchos administradores de sistemas y webmasters muy irresponsables.
- Los encargados de mantener el software seguro sacan actualizaciones cada poco tiempo por algo. No actualizar nuestros sistemas es un grave error.

Como podemos ver, Shodan es un gran buscador, que cada día va avanzando más y más, incluso tenemos la oportunidad de hacer búsquedas de exploits con solo insertar el número de registro de una vulnerabilidad conocida. Si queremos poder hacer infinitas consultas tienemos que pagar solo una vez una pequeña cantidad de dinero, pero merece la pena, y la cantidad de sistemas que podemos ver y los errores de seguridad que podemos llegar a observar navegando un rato por este gran buscador es algo que como conocedor de seguridad informática nos puede enriquecer de una manera increíble.