



Seminario de Tecnología de la Información 2012

Herramientas de Seguridad Informática - Recolección de información

Juan Carlos Miranda

Herramientas de seguridad informática - Recolección de información

Índice de contenido

1 Herramientas de seguridad informática - Recolección de información.....	3
1.1 Objetivos.....	3
1.2 ¿Qué es un pentest?.....	3
1.3 Metodologías para pentest.....	3
2 Recolección de información (Information Gathering).....	4
2.1 Herramientas utilizadas para pentest.....	5
2.2 Footprinting.....	5
2.3 Utilización de páginas públicas para recabar información.....	5
2.4 Utilización de buscadores.....	6
2.4.1 Archivos y su correspondencia.....	6
2.4.2 Comandos del buscador Google.....	7
2.4.3 Comandos del buscador Bing.....	7
2.5 Herramientas especializadas para recolección de datos.....	7
2.5.1 dnsenum.....	7
2.5.2 dnsrecon.....	8
2.5.3 theHarvester.....	8
2.5.4 goofile.....	8
2.5.5 listpurls.....	8
2.5.6 whatweb.....	8
2.5.7 joomscan y plescost.....	8
2.5.8 Anubis.....	8
2.5.9 FOCA.....	9
2.5.10 Maltego.....	10
2.5.11 Resumen de las herramientas presentadas.....	10
2.5.12 Resumen de herramientas según la información a buscar.....	12
3 Conclusión.....	13
4 Bibliografía.....	13



1 Herramientas de seguridad informática - Recolección de información

1.1 Objetivos

- Proveer datos sobre herramientas para recolección de información al momento de realizar un pentest.
- Mostrar ejemplos prácticos sobre el uso de herramientas para recolección de información.

Al finalizar la charla los asistentes contarán con nociones básicas para recabar información para la realización de un pentest utilizando distintas herramientas de seguridad informática.

1.2 ¿Qué es un pentest?

Pentest es un estudio que realizan los profesionales de seguridad para determinar la seguridad en sistemas informáticos y aplicaciones involucradas con los mismos.




Pentest proviene de las letras iniciales de las palabras en inglés “Penetration Test” (Test de penetración).

1.3 Metodologías para pentest


Actualmente para llevar a cabo los pentest se han desarrollado metodologías, las cuales han documentado las diversas pruebas que se pueden realizar en los sistemas computacionales.

No todos los puntos citados en una metodología son utilizados para un pentest, dado que las mismas cubren un amplio espectro de situaciones.

A modo de información se puede citar las siguientes metodologías:

Penetration Testing Execution Standard		http://www.pentest-standard.org
Open Source Security Testing Methodology Manual - OSSTMM		http://www.isecom.org/
Information Systems Security Assessment Framework		http://www.oissg.org/



The Open Web Application Security Project - OWASP		https://www.owasp.org
---	---	---

Habitualmente un pentest contiene las siguientes fases:

- Recolección de información. (Information Gathering)
- Análisis de vulnerabilidades.
- Explotación de vulnerabilidades.
- Generación de informes.

El documento se centra exclusivamente en la fase de “Recolección de información o también denominada Information Gathering”.

2 Recolección de información (Information Gathering)

En el presente documento se desarrollan distintas maneras de recolección de información utilizando fuentes de uso público las cuales son:

- Utilización de páginas públicas para recabar información.
- Utilización de buscadores.
- Herramientas especializadas para recolección de datos.

A las fuentes de uso público se las denomina OSINT (Open Source Intelligence).

Dentro de OSINT se pueden encontrar distintos tipos de áreas para relevar información, las cuales pueden ser:

Empresas/instituciones <ul style="list-style-type: none">• Instalaciones físicas, localización física.• Números de teléfonos.• Organigramas de cargos.• Proveedores.• Empleados.• Sistemas informáticos.
Sistemas informáticos <ul style="list-style-type: none">• Nombres de dominio.• Rangos de Ips públicas.• Servidores DNS.• Servicios de máquinas.• Sistemas operativos.• Metadatos de documentos públicos• Localización geográfica.
Datos de personas <ul style="list-style-type: none">• Números de documentos.• Dirección de correo.• Nombres y apellidos.• Datos en redes sociales (esto puede variar por zona geográfica)



Búsqueda de documentos y metadatos

De los metadatos se busca lo siguiente:

- Usuarios: autores de documentos
- Carpetas donde se crearon los documentos.
- Impresoras.
- Correos.
- Versión de software con el que fue creado.
- Sistemas operativos.
- Software de ofimática.

El éxito de un ingreso a los sistemas depende fuertemente de la información relevada durante la etapa de recolección de datos o “information gathering”.

Los datos recolectados podrán ser utilizados para llevar a cabo pruebas específicas sobre el objetivo elegido.

2.1 Herramientas utilizadas para pentest

Herramientas para realizar tests de penetración existen en gran cantidad, lo importante es saber lo que se desea averiguar del objetivo.

Cuanto más información se recolecte mayor será la probabilidad de éxito.

Las herramientas que se citarán en este documento, es probable que cambien y que sean suplantadas por otras.

Incluso el lector podrá crear su propio software si así lo desea.

En el mercado existen distintas distribuciones Linux, las cuales agrupan un conjunto de herramientas para pentesters.

Para lo tratado en este documento se utiliza Backtrack5.

2.2 Footprinting

Se denomina “footprinting” a las técnicas de recolección de información acerca de sistemas computacionales.

El footprinting puede hacerse en forma pasiva o activa.

- **Footprinting activo:** Es cuando se utilizan herramientas que envían paquetes directamente a los host de la red que se está analizando.
- **Footprinting pasivo:** Es cuando se recolectan datos del objetivo a analizar en forma indirecta.

Según los conocimientos previos sobre los sistemas a analizar se tienen los siguientes tipos de pruebas:

- **Pruebas de caja negra:** ningún conocimiento sobre el sistema a evaluar, es lo más cercano a la visión de un atacante externo.
- **Pruebas de caja blanca:** para saltar la etapa de recolección de información, se proporcionan todos los datos necesarios para que el atacante realice las pruebas directamente en los sistemas.
- **Pruebas de caja gris:** representa un punto intermedio en cuanto a los conocimientos de los sistemas, se provee información al pentester a modo de orientación.

Entre los datos que se pueden averiguar se puede citar:

- Servidores DNS.
- Dominios correspondientes al objetivo seleccionado.
- Rangos de red de servidores públicos.
- Sistemas operativos utilizados.
- Puertos abiertos y los servicios que se ejecutan.
- Archivos publicados en la web.
- Correos electrónicos.

2.3 Utilización de páginas públicas para recabar información

Se pueden utilizar páginas públicas para la recolección de los primeros datos.

Consultas de dominio utilizando http://nic.com	Búsqueda de informaciones tales
---	---------------------------------



herramientas web	http://nic.py http://whois.domaintools.com/ http://www.robtex.com http://mxtoolbox.com http://www.netcraft.com	como: DNS, Blacklists, dominios de correo.
	http://www.cuwhois.com	Análisis de enlaces de la página. Análisis de cabeceras. Análisis de metatags. Información sobre servidores activos.

2.4 Utilización de buscadores

Se puede utilizar los buscadores para encontrar información acerca de un objetivo específico.

Entre los datos que podemos recabar en los buscadores se tiene:

- Teléfonos.
- Cuentas de correos.
- Dominios relacionados a los sitios.
- Archivos de distintos tipos (.doc, .xls, .pdf, .pps, jpeg, .png, etc..) para obtención de metadatos.
- Dominios relacionados con sitios.

Para el caso de los buscadores Bing y Google, se permiten opciones avanzadas, las cuales pueden incluir operadores lógicos del tipo: AND, OR.

Utilizando los buscadores es posible conocer datos de un sitio web, los cuales no se encuentran accesibles desde la página principal.

2.4.1 Archivos y su correspondencia

A continuación, se exponen extensiones de archivos que pueden ser incluidas en las búsquedas para recolección de datos.

Extensión		Extensión	
csv	Texto separado por comas	pptx	Presentación Microsoft Power Point 207/2010 XML
doc	Documento de texto Microsoft Word 97/2000/XP/2003	rtf	Documento de texto Rich Text Format
docx	Documento de texto Microsoft Word 2007/2010 XML	svg	Scalar Vector Graphics
ica	Archivo relacionado con productos Citrix	svgz	Compressed Scalable Vector Graphics
indd	Adobe InDesign	sxc	Hoja de Cálculo OpenOffice.org 1.0
odg	Dibujo Impress	sxd	Presentación OpenOffice.org 1.0
odp	Presentación	sxi	Presentación OpenOffice.org 1.0
ods	Hoja de Cálculo OpenOffice.org	txt	Texto ASCII
odt	Documento de texto OpenOffice.org	xls	Microsoft Excel 97/2000/XP/2003
pdf	Documento Acrobat Reader	xlsx	Microsoft Excel 2007/2010 XML
pps	Presentación Microsoft Power Point 97/2000/XP/2003 Autoplay	jpg	Joint Photographic Experts Group, es el formato de imágenes más utilizado en Internet.



ppsx	Presentación Microsoft Power Point 207/2010 XML Autoplay	png	Portable Network Graphics.
ppt	Presentación Microsoft Power Point 97/2000/XP/2003	gif	Formato ampliamente utilizado en Internet.

2.4.2 Comandos del buscador Google

Ayuda sobre el buscador Google http://www.elhacker.net/trucos_google.html

También se pueden buscar los verbos bajo la frase “Google Dorks”.

Verbos	Función
site:	Búsqueda dentro de un dominio específico.
filetype:, ext:	Búsqueda de archivos en un formato determinado.
intitle:	Buscar páginas con palabras en el campo title de las páginas.
inurl:	Buscar páginas con palabras específicas en la URL.
Ejemplos	
Búsqueda de documentos .DOC	filetype:doc site: midominio.com
Búsqueda de cuentas de correos	@midominio.com site: midominio.com.py
Búsqueda de teléfonos en páginas en español	Tel. site:midominio.com.py
Buscar índices en sitios web	intitle:index of site:uci.edu.py

2.4.3 Comandos del buscador Bing

Ayuda del buscador Bing <http://onlinehelp.microsoft.com/es-es/bing/ff808421.aspx>

Verbos	Función
filetype:	Búsqueda de archivos según la extensión.
inanchor:, inbody:, intitle:	Devuelve resultados con las palabras que se encuentran en el enlace, cuerpo del documento, el título.
site:	Devuelve las páginas web pertenecientes al sitio que se especifica.
IP:	Busca datos relacionados con un número de IP específica.
Ejemplos	
Búsqueda de documentos .PDF	filetype:pdf site: midominio.com
Búsqueda de cuentas de correos	@midominio.com site:midominio.com
Búsqueda de teléfonos en páginas en español	Tel. site:midominio.com.py

2.5 Herramientas especializadas para recolección de datos

2.5.1 dnsenum

Es una herramienta que permite la enumeración de hosts, servidores dns de un dominio específico, servidores de correos relacionados al dominio.



Se utiliza desde la línea de comandos y se encuentra incluida en la distribución Backtrack5 bajo el directorio /pentest/enumeration/dns/dnsenum. Ej: dnsenum.pl dominio.com

2.5.2 dnsrecon

Es otra herramienta que permite obtener datos sobre los DNS correspondientes a un dominio, incluye características similares a dnsenum.

Se utiliza desde la línea de comandos y se encuentra incluida en la distribución Backtrack5 bajo el directorio /pentest/enumeration/dns/dnsrecon. Ej: dnsrecon.pl -d dominio.com

2.5.3 theHarvester

Es una herramienta que permite listar las cuentas de correos de un dominio utilizando los motores de búsqueda Google, Bing, Linkeding.

Permite obtener información sobre host, direcciones IP que se encuentran asociadas al dominio.

Se utiliza desde la línea de comandos y se encuentra incluida en la distribución Backtrack5 bajo el directorio /pentest/enumeration/theharvester. Ej: theHarvester.py -d dominio.com -l 500 -b google

2.5.4 goofile

Es una herramienta que permite listar archivos de distintos tipos dentro de un dominio, utilizando el motor de búsqueda Google.

Se utiliza desde la línea de comandos y se encuentra incluida en la distribución Backtrack5 bajo el directorio /pentest/enumeration/google/goofile. Ej: goofile.py -d dominio.com -f pdf.

2.5.5 listpurls

Herramienta que permite obtener los enlaces de una página web, incluyendo correos que se encuentren como contenidos. Es útil para estudiar la arquitectura del sitio web.

Se utiliza desde la línea de comandos y se encuentra incluida en la distribución Backtrack5 bajo el directorio /pentest/enumeration/list-urls. Ej: /listpurls.py -l www.dominio.com

2.5.6 whatweb

Herramienta que permite obtener datos sobre las tecnologías utilizadas en un sitio web. Sirve para relevar datos de los sitios.

Se utiliza desde la línea de comandos y se encuentra incluida en la distribución Backtrack5 bajo el directorio /pentest/enumeration/web/whatweb. Ej: ./whatweb www.dominio.com

2.5.7 joomscan y plescost

Herramientas incluidas en Backtrack5 en el directorio /pentest/web/scanners/ para la verificación de vulnerabilidades en gestores de contenidos web.

Joomscan permite analizar frameworks Joomla y plescost para Wordpress.

2.5.8 Anubis

Es una herramienta creada para la recolección de datos de fuentes públicas. Permite realizar búsquedas utilizando



motores buscadores conocidos como Google y Bing, esta es su mayor bondad.

Cuenta con características que permiten la recolección de datos bajo diferentes métodos.

Es una gran ayuda al momento de realizar consultas dado que presenta varios tipos de consultas bien definidas y organizadas.

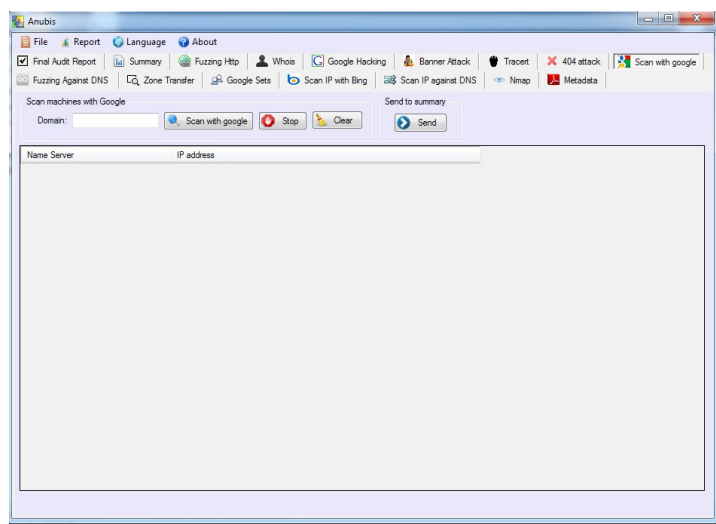


Ilustración 1: Anubis

2.5.9 FOCA

FOCA es una herramienta que combina muchas funciones. La principal función es otorgar al usuario los datos acerca de un dominio específico.

Entre los datos que permite visualizar están:

- Rangos de IPs.
- Nombres de dominios y subdominios.
- Banner fingerprinting.
- Archivos con posibles metadatos.

Es una herramienta mas que completa al momento de relevar datos.

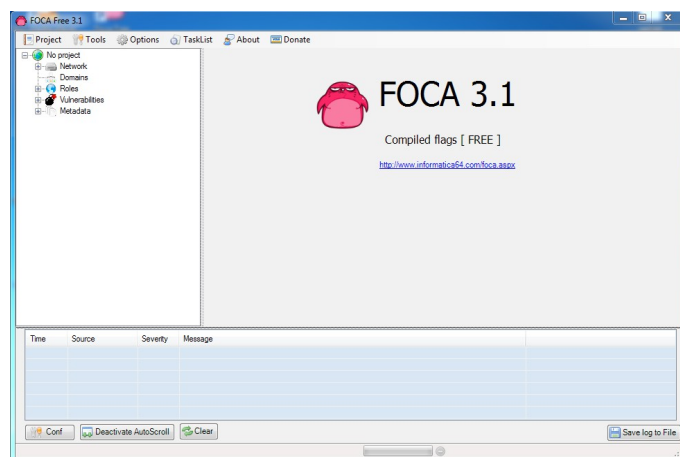


Ilustración 2: Foca

2.5.10 Maltego

Maltego es una herramienta que permite recolectar información de fuentes públicas y puede ser utilizada para determinar relaciones y enlaces entre: personas, grupos de personas, compañías, organizaciones, sitios web, documentos, datos de infraestructura en Internet (DNS, redes, direcciones IP).

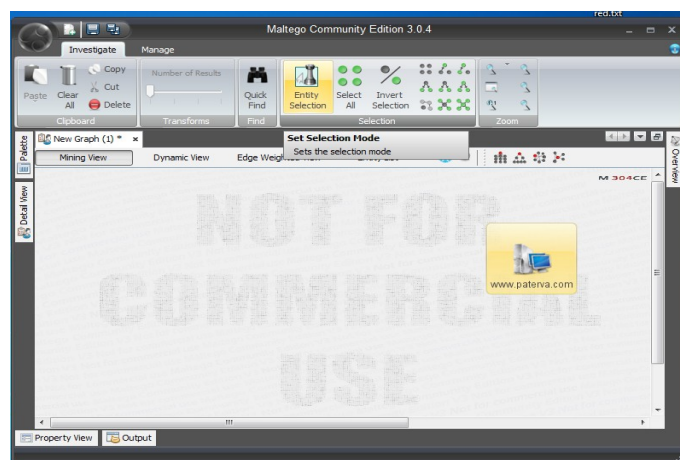


Ilustración 3: Maltego

2.5.11 Resumen de las herramientas presentadas

En la fase de “Recolección de datos” se utilizan varias herramientas, las mismas se presentan a continuación:

Nombre	Función	Autor/es	URL	Disponibilidad (pago/gratuito)	S.O recomendado para ejecutarlo
dnsenum	Enumerar datos relacionados al DNS de un dominio dado.	Filip Waeytens filip.waeytens@gmail.com ail.com		Gratuito, disponible en Backtrack5	Linux



	Permite conocer los servidores de correo relacionados al dominio. Script Perl	tix tixxDZ tixxdz@gmail.com			
theHarvester	Recolección de correos electrónicos en buscadores. Script Python.	Christian Martorella cmartorella@edge-security.com	Http://code.google.com/p/theharvester	Gratuito, disponible en Backtrack5	Linux
goofile	Búsqueda de documentos utilizando google. Script Python.	Tomas Richards www.g13net.com	Http://code.google.com/p/goofile	Gratuito, disponible en Backtrack5	Linux
list-urls	Permite listar urls de una página web. Script Python.	Mark Pilgrim mark@diveintopython.org	Http://diveintopython.org	Gratuito, disponible en Backtrack5	Linux
whatweb	Reconocimiento de las tecnologías utilizadas en un sitio web, correos electrónicos y otros datos. Desarrollado en Ruby	Andrew Horton	Http://www.morningstarsecurity.com/research/whatweb	Gratuito, disponible en Backtrack5	Linux
joomscan	Identificación de vulnerabilidades para Joomla. Script python.	Aung Khant	Http://yehg.net/lab	Gratuito, disponible en Backtrack5	Linux
plesco	Identificación de vulnerabilidades para Wordpress	Francisco J. Gomez. Daniel Garcia Garcia	Http://iniqua.com/labs	Gratuito, disponible en Backtrack5	Linux
Anubis	Herramienta que permite distintos tipos de test para la realización de recolección de información.	Juan Antonio Calles García	Http://elblogdecalls.blogspot.com http://flu-project.com	Gratuito	Windows
FOCA	Combina muchas funcionalidades en una sola herramienta, desde reconocimiento de hosts hasta metadatos de documentos. Es una de las herramientas más completas. Se encuentra disponible para entornos Windows.	Informática 64	Http://www.informatica64.com	Versiones free y paga.	Windows
Maltego	Recolecta datos basados en fuentes	Paterva	Http://www.paterva.com	Versión comercial.	Linux/ Windows



	públicas, dibuja las relaciones entre los objetos encontrados.			Versión no comercial (descarga gratuita)	
--	--	--	--	--	--

2.5.12 Resumen de herramientas según la información a buscar

Es difícil realizar una clasificación clara para las herramientas presentadas, debido a las funciones que tienen. También es necesario hacer notar que un mismo dato puede ser obtenido por varias herramientas.

Siempre se deberá contrastar la información obtenida contra otras herramientas.

La siguiente tabla resume las herramientas que pueden utilizarse según los datos expuestos al comienzo del documento.

Datos a recolectar	Fuentes y herramientas
Empresas/instituciones Instalaciones físicas, localización física. Números de teléfonos. Organigramas de cargos. Proveedores. Empleados. Sistemas informáticos.	Buscadores Google, Bing. Folletería de marketing. Cartelería. Trashing (recolección de residuos) Consultas telefónicas. Anubis. Foca. Maltego.
Sistemas informáticos <ul style="list-style-type: none">Nombres de dominio.Rangos de Ips públicas.Servidores DNS.Servicios de máquinas.Sistemas operativos.Metadatos de documentos públicosLocalización geográfica.	Buscadores Google, Bing Dnseum theHarvester goofile. Anubis.
Datos de personas <ul style="list-style-type: none">Números de documentos.Dirección de correo.Nombres y apellidos.Datos en redes sociales (esto puede variar por zona geográfica)	Redes sociales. Buscadores Google, Bing. Redes sociales. theHarvester list-urls whatweb joomscan plesco
Búsqueda de documentos y metadatos <ul style="list-style-type: none">De los metadatos se busca lo siguiente:Usuarios: autores de documentosCarpetas donde se crearon los documentos.Impresoras.Correos.Versión de software con el que fue creado.Sistemas operativos.Software de ofimática.	FOCA



3 Conclusión

Se han presentado diversas herramientas para recabar datos, pero al momento de registrar los datos obtenidos será necesario hacerlo en forma ordenada y metódica para poder aprovecharlos al máximo.

“No solo es necesario tener la información, sino saber encontrarla y utilizarla cuando se necesite”.

Juan Carlos Miranda.

juancarlosmiranda81@gmail.com

4 Bibliografía

- Google Guide. <http://www.googleguide.com> [25/09/2012]
- Bing Help, “Opciones de búsqueda avanzadas”. <http://onlinehelp.microsoft.com/es-es/bing/ff808421.aspx> [25/09/2012.]
- elhacker.net, “Trucos Google, Trucos de búsqueda y Curiosidades sobre el buscador”. http://www.elhacker.net/trucos_google.html, [25/09/2012]
- INTECO CERT, “Pentest: Recolección de Información (Information Gathering)”. http://cert.inteco.es/extfrontinteco/img/File/intecocert/EstudiosInformes/cert_inf_seguridad_information_gathering.pdf [02/09/2012]
- Juan Antonio Calles García, Pablo González Pérez, “La Biblia del Footprinting”. http://www.flu-project.com/descargasDirectas/pdf/La_Biblia_del_Footprinting.pdf [02/09/2012]