REVISTA DIGITAL DE LAS TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES
TELEM@TICA
Revista Telem@tica. Vol. 13. No. 1, enero-abril, 2014, p. 20-30 ISSN 1729-3804

SISTEMA PARA LA GESTION DE ENLACES O URL MALICIOSOS
MANAGEMENT SYSTEM FOR LINKS OR MALICIOUS URL

Ing. Daniel Ramón Torres Miyares 1, MsC. Nelson William Gamazo Sánchez 2
1 Segurmatica, Cuba, drtorres@segurmatica.cu, Zanja No. 651 Esq. a Soledad. Centro Habana. Ciudad de la Habana.
2 Segurmatica, Cuba, ngamazo@segurmatica.cu

RESUMEN: El hospedaje de programas malignos en Internet ha aumentado considerablemente en los últimos años. Como resultado es posible obtener, por diferentes vías, millones de enlaces o URLs desde los cuales se puede descargar programas malignos. Muchos de estos enlaces tienen la característica de ser dinámicos, por ejemplo, enlaces a un mismo programa maligno pueden desaparecer y reaparecer como otro enlace diferente, en ocasiones se activan o están disponibles durante un corto periodo de tiempo. Debido a esto, la implementación de un sistema capaz de encontrar, descargar con priorización y clasificar dichos enlaces permitiría obtener una base de datos de muestras de programas malignos actualizados, además de una lista de URLs activos en todo momento. El sistema expuesto en este trabajo es capaz de encontrar, descargar y dar seguimiento a los URLs que apuntan a programas malignos. El sistema además permite detectar nuevos repositorios de programas malignos a partir de los enlaces previamente conocidos.

Palabras Clave: Programas malignos, capturar, seguimiento, enlaces, antivirus, Cuba.

ABSTRACT: *The hosting of malicious programs on the Internet has increased considerably in recent years. As a result it is possible to obtain, in different ways, millions of links or URLs from which you can download malicious software. Many of these links have the characteristic of being dynamic, for example, links to the same malignant program can disappear and reappear as another link different, sometimes activated or available for a short period of time. Because of this, the implementation of a system capable of finding, downloading with prioritization and classify such links, would provide a database of current malware samples, and a list of URLs active at all times. The system described in this paper is able to find, download and track the URLs pointing to malicious programs. The system also allows detecting new malicious software repositories from previously known links.*

Key Words: Malware, detention, monitoring, links, antivirus, Cuba.

INTRODUCCIÓN

En la actualidad se ha propiciado un aumento significativo en el hospedaje de programas malignos en la Web. En los últimos años la mayor parte del malware ha sido creado con un fin económico o para obtener beneficios en algún sentido. Esto es debido a la decisión de los autores de malware de sacar partido monetario de los sistemas infectados.

En la empresa Segurmática donde se desarrolla el Segurmatica Antivirus (Segav), se necesita mantener una base de muestras de programas malignos que contenga muestras que reflejen, con la mayor actualización posible, las muestras que se generan diariamente en Internet y que pueden afectar los usuarios. Esta base de datos es necesaria para el trabajo de identificación y desinfección del motor del antivirus.

De aquí surge la necesidad de implementar un sistema capaz de localizar enlaces de Internet con programas malignos, descargar estas muestras de Internet, procesarlas, clasificarlas y almacenar estas muestras para mejorar el trabajo de las personas involucradas en el análisis de programas malignos.

El objetivo del presente trabajo es dar a conocer el uso de un sistema capaz de localizar enlaces de Internet con programas malignos, el cual va a posibilitar encontrar repositorios de programas malignos donde se hospeden estos. También va a permitir crear listas de dominios y direcciones malignas en Internet, las cuales pueden ser utilizadas como listas de bloqueos de navegación aumentando la seguridad en las empresas cubanas ante ataques manejados por descargas los que han aumentado en los últimos años significativamente. [1] [2]

ARQUITECTURA Y FUNCIONAMIENTO DEL SISTEMA

El sistema está compuesto por diferentes módulos que desempeñan diferentes actividades:

•Servicio Web: Es el encargado de la gestión de descarga de enlaces de internet brin-dando funcionalidades para ello.

• Interfaz Web: Página web que permite interactuar con los resultados del sistema y monitorear el estado de las etapas de obtención, clasificación y almacenamiento de muestras. Permite, también, insertar e identificar nuevos enlaces manualmente. Contiene estadísticas del sistema.

• Programador de tareas: Controlador de las actividades de las distintas fases que se realizan programadas en el sistema.

• Reportes: Servidor de reportes generados a partir de los resultados de procedimientos almacenados de la base de datos.

• Base de datos: Servidor de base de datos donde se almacenan todos los datos correspondientes a los enlaces, descargas, clasificaciones, entre otros.

• Líneas de comandos de antivirus: Conjunto de líneas de comandos de los antivirus utilizados en la identificación y clasificación de muestras, entre los que se tiene a Kaspersky Antivirus, Bitdefender Antivirus, entre otros. [3]

La interacción de estos módulos se ve reflejada en la Figura. 1 que muestra a los componentes del sistema.

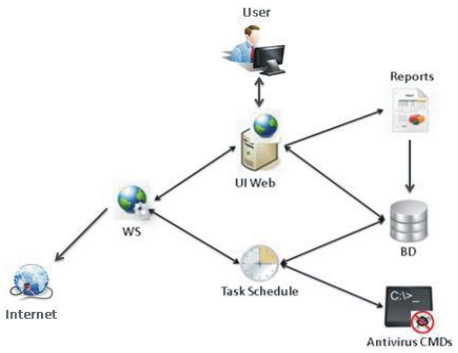


Figura 1. Arquitectura del sistema.

El trabajo del sistema se realiza automático, ya que el programador de tareas ejecuta las diferentes actividades del sistema durante determinado tiempo. El primer paso comienza por la invocación del recolector de enlaces, el cual utiliza varios métodos de recolección que se expondrán en el punto 2.1. Tras la recolección, se ejecuta la obtención de enlaces, para lo que se utilizan las funcionalidades del servicio Web de descarga para obtener los enlaces que necesitan ser descargados. La próxima actividad a ejecutarse es la comprobación de las muestras descargadas por las líneas de comandos de antivirus para catalogar los archivos; en dependencia del resultado obtenido por los antivirus, se almacenan las muestras.

La interfaz Web muestra los reportes que se generan en el servidor de reportes, basados en los procedimientos almacenados de la base de datos. Esta también le brinda información directa a la interfaz Web referente a los últimos enlaces obtenidos.

El sistema cuenta con varias fases: la recolección de enlaces, la descarga de las muestras, el procesamiento de los archivos descargados, la clasificación y almacenamiento de estos y la generación automática de estadísticas del sistema para comprobar su funcionamiento.

Obtención de enlaces

Para la captura de enlaces se utilizan varios mecanismos los cuales permiten al sistema recolectar enlaces automáticos diarios y programadamente

Sitios

En la Web existen empresas o grupos que publican, gratuitamente, enlaces a programas malignos como son los casos de MalwareDomainList, MalwareBlackList, ThreatExpert, Malc0de, entre otros.

Para obtener estos enlaces de los sitios de publicación, el sistema tiene almacenadas las estructuras HTML2 [6] de estos sitios para poder navegarlas de forma automatizada, buscando obtener todos los enlaces publicados por estos en sus páginas, como se muestra en la Figura. 2.

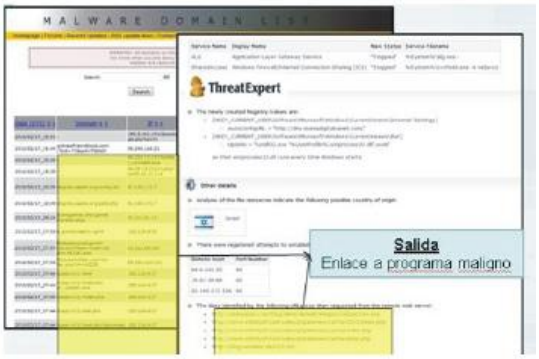


Figura 2. Extracción de enlaces.

Spam

El correo masivo supone actualmente la mayor parte de los mensajes electrónicos intercambiados en Internet. Los atacantes utilizan los correos spam para enviar enlaces maliciosos enmascarados con textos de promoción. [4]

La publicación de direcciones de correo en sitios Web y fórums maliciosos propicia que se añadan las direcciones de correo del sistema de captura de spam a las listas de distribución de correos utilizadas por los generadores de correos masivos, aumentando los spam capturados. Los correos recibidos son analizados buscando enlaces para adicionar al sistema. En la Figura 3 se muestra un ejemplo de esto.

Figura. 3 Ejemplo de Spam con malware.

Otras empresas

Se recolectan enlaces también de otras empresas que los proveen gratuitamente por convenios de cooperación o suscripción en listas de correos. Los enlaces se reciben por mensajes de correo que se reciben en **urlsamples@segurmatica.cu** y se organizan en subcarpetas según la fuente, como se puede observar en la Figura 4. Constantemente el sistema revisa los buzones buscando nuevos correos para tomar los enlaces e insertarlos en la base de datos para ser descargados posteriormente. Uno de los grupos de donde se obtienen enlaces es Viruswatch el que envía su lista de enlaces tras la suscripción a su página Web.

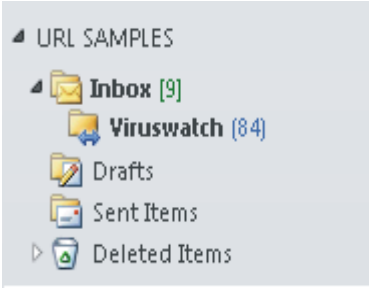


Figura 4 Recepción de enlaces por correo.

Búsqueda de sitios con lista de archivos.

Una de las estrategias para obtener enlaces se basa en que un pequeño porcentaje de los servidores dedicados a publicar programas malignos permiten explorar los directorios de archivos mostrando listas de contenidos y carpetas. Aprovechando los enlaces infestados que se tienen en la base de datos se comprueba a los directorios raíces para cada uno de los diversos enlaces. A estos directorios se les analiza la respuesta del servidor para determinar si muestran listas de archivos o no. [5] Para comprobar si un directorio es explorable, se analiza la estructura HTML de la respuesta web comparándola con las estructuras conocidas de exploración de directorios de los principales servidores webs como IIS[3], Apache y Nginx. En caso de que el servidor permita la exploración de archivos se extraen de forma recursiva los enlaces que se encuentran en todos los directorios del servidor. [7][8][9] Los enlaces obtenidos son insertados en el ciclo del sistema para ser descargados y analizados.

Este algoritmo permite detectar repositorios de programas malignos como es el caso que se observa en la Figura. 5, existiendo repositorios en donde se actualizan nuevas versiones de muestras en las que trabajan los desarrolladores, lo que permite tener muestras al día.

Figura 5. Repositorio de programas malignos.

Generación de nuevos enlaces a partir de la información de los ya existentes

Otro algoritmo que ha aportado enlaces nuevos al sistema se basa en buscar servidores duplicados los cuales contengan las mismas muestras pues es muy común que los desarrolladores de programas malignos pongan servidores duplicados en diferentes direcciones de Internet. Para realizar esta operación se buscan similitudes entre servidores; es decir, que contengan más de n muestras en común

con caminos iguales hasta el archivo. Esto se les aplica a los enlaces que se tienen almacenados. Al encontrar estas posibilidades se les asignan direcciones de un servidor al otro y viceversa, buscando la existencia de nuevos enlaces.

Este algoritmo ha permitido encontrar servidores malignos de internet con las mismas estructuras de archivos, demostrando que los atacantes despliegan diferentes servidores para tener una menor posibilidad de que les sean bloqueadas sus direcciones malignas en la web.

Descarga de muestras

Los nuevos enlaces almacenados en la base de datos se envían a un servicio Web que se encarga de descargarlos, obtener los datos referentes a la dirección IP, localizar los enlaces y generar un reporte con esta información que los actualiza en la base de datos.

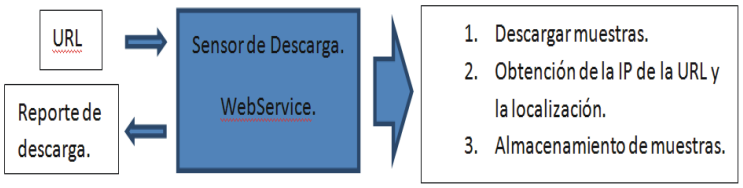


Figura 6. Diagrama de interacción con el servicio web.

Con el objetivo de comprobar la disponibilidad de las muestras descargadas y si han sufrido algún cambio por parte de los desarrolladores se encuentran hospedados varios sensores que permite descargarlos nuevamente, estos archivos y actualizar la clasificación y datos.

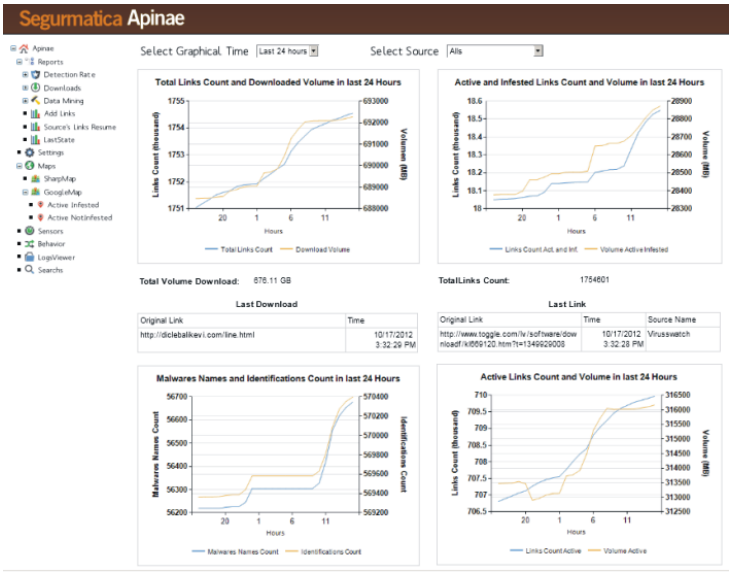
Clasificación y almacenamiento

Las muestras descargadas se analizan con líneas de comandos de los antivirus más utilizados en la actualidad, como Kaspersky y BitDefender. También se analizan con Segurmática Antivirus para poder clasificar los archivos en contaminados o no dándole prioridad de trabajo a los que no son detectados por Segurmática Antivirus y si por otros, buscando obtener mejores resultados con el antivirus de esta empresa.

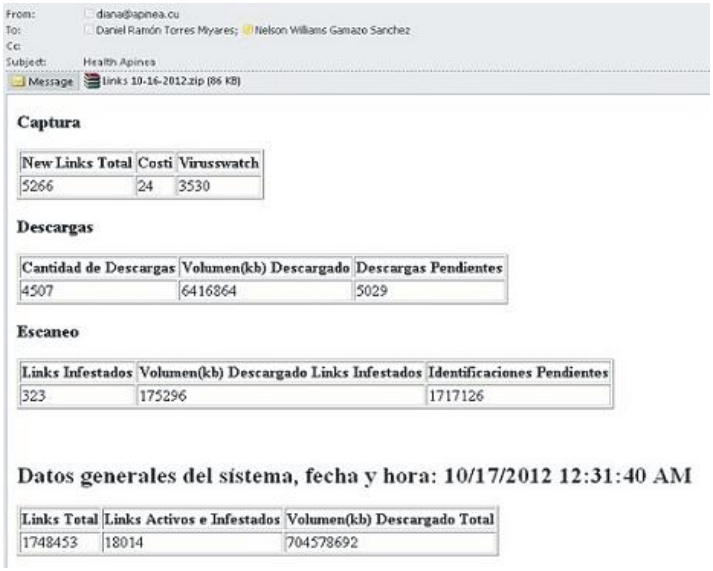
Las muestras se almacenan según sean programas malignos o no y del tipo de archivo que sea binario o texto, utilizando una estructura de carpetas basadas en el hash del archivo para poder obtenerlo fácilmente.

Generación de estadísticas

El sistema contiene actividades que reportan el funcionamiento del sistema. Entre ellas se tiene gráficas del comportamiento de cada una de las fases del sistema, como se puede apreciar en la Figura 7, el sistema incluye estadísticas de rendimiento, de la captura de enlaces, de los sensores de descarga del análisis de muestras, entre otros.



Otra forma de controlar el funcionamiento es mediante alertas de correo electrónico diarias, las cuales contienen resultados generales del comportamiento de la aplicación, como se observa en la Figura 8, y estadísticas de cada una de las fuentes de las que se obtienen los enlaces.



From: diana@apinea.cu
To: Daniel Ramón Torres Miyares; Nelson Williams Gamazo Sanchez
Cc:
Subject: Health Apinea
Message: Links 10-16-2012.zip (86 KB)

Captura

New Links Total	Costi	Virusswatch
5266	24	3530

Descargas

Cantidad de Descargas	Volumen(kb) Descargado	Descargas Pendientes
4507	6416864	5029

Escaneo

Links Infestados	Volumen(kb) Descargado Links Infestados	Identificaciones Pendientes
323	175296	1717126

Datos generales del sistema, fecha y hora: 10/17/2012 12:31:40 AM

Links Total	Links Activos e Infestados	Volumen(kb) Descargado Total
1748453	18014	704578692

Figura 8. Alerta de correo.

RESULTADOS OBTENIDOS

En los dos años de explotación del sistema en Segurmatica se han obtenido resultados entre los que se destaca la captura de 176,0384 enlaces con la utilización de los diversos mecanismos creados. De estos, 162,310 enlaces contienen programas malignos identificados por los sistemas antivirus. Entre los enlaces capturados 715,844 están activos en la actualidad a los cuales se pueden acceder en internet y de estos 19,200 están infestados.

El sistema ha descargado un total de 682.410GB que representan 275,1974 archivos ya que hay enla-ces que contienen comprimidos, de los cuales 362,262 contienen programas malignos. Estos datos se ven reflejados en la Figura 9.

State	Bin	Text	Total	Size
Links	258101	398218	1760384	682.41 GB
Links With Malware	73052	42073	162310	238.62 GB
Links Active	172156	286029	715844	315.25 GB
Links Active with Malware	14691	4380	19200	197.06 GB
Files Downloaded	446904	1295946	2751974	
Files Extracted	143023	87385	362262	
Files Extracted with Malware	143023	87385	362262	

Figura 9. Resultados del sistema de captura.

El sistema brinda la posibilidad de hacer un estudio de los principales malware publicados en Internet, como se puede observar en la Figura. 10. Esto tiene gran importancia para los analistas, ya que les permite centrarse en los programas malignos más distribuidos en Internet. También permite ver los principales países donde se publican malware en el mundo. En la Figura 11 se ve la utilización de un servicio de mapas para mostrar la cantidad de programas malignos detectados por países, lo que permite determinar los dominios y direcciones IP que representan una amenaza.

Malware Detection Rate

Binary Files Malware names

Malware Name	Files Detected
Trojware.Trojan-Downloader.Win32.Lipler.axkd	62228
Gen:Heur.NaviPromo.5	40708
Trojan.Skintrim.6043	6287
Adware not-a-virus:AdWare.Win32.iBryte.x	5324
Adware not-a-virus:AdWare.Win32.ScreenSaver.e	4254
DeepScan:Generic.Skintrim.4C4A9370	4144
Virware.HEUR:Trojan.Win32.Generic	3188
Gen:Variant.Adware.Hotbar.1	2441
Generic.Adw.ScreenBlaz.39F7DB82	2179
Troyano.W32.DELF.UVK	1851

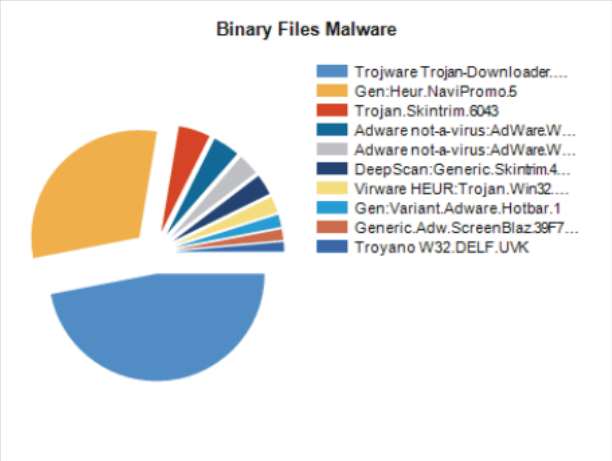


Figura 10. Programas malignos detectados.

Figura 11. Sistema de mapas.

LIMITACIONES

Los servidores de programas malignos en algunos casos están dirigidos a países y zonas específicas y bloquean el acceso a estos desde otros países. Esto se pueden evitar creando sensores de descarga en diferentes países en cooperación con otras empresas que desarrollan antivirus.

Las comprobaciones de enlaces necesitan la descarga de Webs para buscar enlaces y con un ancho de banda limitado, esta operación se ve afectada.

El servidor de base de datos, por las constantes operaciones que realiza no cumple con las necesidades del sistema, por lo que se necesitaría un servidor con altas prestaciones.

CONCLUSIONES

El presente trabajo demuestra que es posible implementar un sistema de captura de muestras de programas malignos en internet, el cual ha permitido la obtención de muestras en la empresa Segurmática, brindando la posibilidad al grupo de análisis de contar con mayor número de muestras y estadísticas de estas para optimizar su trabajo. El ciclo de funcionalidades del sistema se realiza de forma automática, con alertas ante errores logrando funcionar sin necesidad de control manual de este.

Con las estadísticas de los datos obtenidos se puede determinar dominios maliciosos así como direcciones IP en donde se alojan programas malignos. Con los enlaces obtenidos se pueden crear listas de bloqueos tanto de enlaces como para dominios, Estas listas son muy utilizadas en la actualidad ya que aumentan la seguridad en la navegación a los usuarios de la Web.

Con la captura de enlaces se le puede dar seguimiento a enlaces que pertenezcan a dominios cubanos, detectando así dominios comprometidos por atacantes externos con el objetivo de publicar programas malignos en Cuba.

REFERENCIAS

1. Sophos Ltd. and Sophos Group: "Sophos Security Threat Report 2012", Boston, USA | Oxford, UK, 2012.

2. Cisco Systems, Inc.: "Cisco 1Q11 Global Threat Report", San José, CA, 2011.

3. Tomas Arlt, Andreas Clementi, Philippe Rodlach and Peter Stelzhammer: "Review of IT Security Suites for Corporate Users, 2010", Nor-derstedt, Germany, 2010.

4. Bruce C. Brown: "How to Stop E-mail Spam, Spyware, Malware, Computer Viruses and Hackers", Ocala, Florida, pp. 29-33, 2011.

5. Dafydd Stuttard and Marcus Pinto: "The Web Application Hacker's Handbook: Finding and Exploiting Security Flaws", Indianapolis, Indiana, Chapter 18: Vulnerable Server Configuration, 2011.

6. Gary B. Shelly and Denise M. Woods: "HTML: Introductory Concepts and Techniques", Boston, USA, 2009.

7. Kenneth Schaefer, Jeff Cochran, Scott Forsyth, Rob Baugh, Mike Everest and Dennis Glendenning: "Professional IIS 7", Chapter 17, March 2008.

8. The Apache Software Foundation: "Apache HTTP Server 2.2 Official Documentation", 2010.

9. Dipankar Sarkar: "Nginx 7 Web Server Implementation Cookbook", Chapter 1 pp. 48-50 UK, May 2011.