Universidad Internacional de La Rioja
Máster universitario en Seguridad Informática

Detector de malware mediante funciones hash criptográficas en infraestructuras multiplataforma descentralizadas

Trabajo Fin de Máster

presentado por: Rodríguez Galiano, David

Director/a: Muñoz Muñoz, Alfonso

Ciudad: Madrid
Fecha: 27 de agosto de 2015

Resumen

Vivimos en una sociedad cada vez más globalizada e interconectada, donde los datos continuamente se intercambian y crecen de forma exponencial.

Junto a las continuas mejoras tecnológicas para dar servicio a todas las demandas, también crece el número de elementos que pueden infringir las políticas de seguridad y consiguen acceder y modificar información sensible o confidencial, pudiendo dañar nuestra fuente de negocio.

Usualmente, con el fin de conseguir dichos objetivos, se utilizan programas conocidos de forma genérica como *malware* que se infiltran en los equipos y proporcionan una gran capacidad de detección y explotación de las vulnerabilidades de los sistemas.

Internet cada día está más infectada, por lo que es requerimiento indispensable disponer de elementos que protejan todo el parque tecnológico de empresas y organizaciones.

En este trabajo se pretende mostrar una forma en la que cualquier organización, independientemente de su naturaleza y tamaño, puede detectar archivos maliciosos (virus, gusanos, troyanos, rootkits, spyware, adware, etc.) de forma ágil, gratuita y confiable, mediante funciones hash criptográficas.

Palabras Clave: hash, sha256, antivirus, malware, VirusTotal

Abstract

We live in an increasingly globalized and interconnected society, where data is exchanged continuously and grows exponentially.

Along with the continuous growing technology improvements to service all the needs, so does the number of items that may violate security policies and manage to access and modify sensitive or confidential information that may damage our business.

Usually, in order to achieve these objectives, there are used programs as malware that infiltrate computers and are capable to detect and use the vulnerabilities found in the systems.

Internet is becoming increasingly infected, so it is an indispensable requirement to have elements to protect all the technological park of businesses and organizations.

This paper aims to show a way in which any organization, regardless of its type or size, can detect malicious files (viruses, worms, Trojans, rootkits, spyware, adware, etc.) in a fast, free and reliable way through cryptographic hash functions.

Keywords: hash, sha256, antivirus, malware, VirusTotal

Índice de contenido

Índice de contenido 4

Índice de ilustraciones 6

Índice de tablas 8

Capítulo 1 Introducción 9

Capítulo 2 Objetivos 11

2.1 Objetivo general 11

2.2 Objetivos específicos 12

Capítulo 3 Metodología 13

3.1 Firmas y funciones hash criptográficas 13

3.2 Proyecto VirusTotal 15

3.3 Estructura 16

3.3.1 Cliente 17

3.3.2 Sistema gestor de base de datos MySQL 17

3.3.3 Servidor web 19

Capítulo 4 Flujo de datos 20

4.1 JSON 20

4.2 EICAR 20

4.3 Flujo general 21

4.3.1 Flujograma 21

Capítulo 5 Ejecución y resultados 23

5.1 Requisitos 23

5.2 Parametrización e invocación 25

5.3 Resultados....29
5.4 Almacenamiento de los resultados en la base de datos local....33
5.5 Visualización vía web....37
Capítulo 6 Modelados de datos y análisis de antivirus....40
6.1 Número total de ficheros analizados....40
6.2 Número de positivos detectados por antivirus....44
6.3 Relación de falsos negativos....48
Capítulo 7 Incidencias en el uso del servicio VirusTotal....54
Capítulo 8 Objetivos logrados y trabajo futuro....56
Referencias bibliográficas....58
Anexo A – Código fuente y licenciamiento....60

Índice de ilustraciones

Ilustración 1: Estructura general 12

Ilustración 2: Estructura particular 16

Ilustración 3: Diagrama entidad-relación 19

Ilustración 4: Flujograma detector de malware 22

Ilustración 5: Área privada y solicitud de licencia a VirusTotal 25

Ilustración 6: Ejemplo de configuración de parámetros de conexión (*VTMySQL.pm*) 26

Ilustración 7: Cambio de ubicación de las librerías del detector de malware (*search.pl*) 26

Ilustración 8: Ejemplo de ejecución de análisis de un fichero 27

Ilustración 9: Ejemplo de ejecución de análisis de un directorio 27

Ilustración 10: Análisis de un fichero no tratado previamente 28

Ilustración 11: Análisis de un fichero tratado previamente 28

Ilustración 12: Petición encolada en VirusTotal 29

Ilustración 13: Relación de ficheros subidos a *VirusTotal* mediante la API 32

Ilustración 14: Detalles de consumo de la licencia de VirusTotal 33

Ilustración 15: Almacenamiento de datos en la tabla "*malware*" 34

Ilustración 16: Almacenamiento de datos en la tabla "*machine*" 36

Ilustración 17: Almacenamiento de datos en la tabla "*antivirus*" 37

Ilustración 18: Protección de plataforma web para detector de malware 38

Ilustración 19: Formulario inicial de búsqueda de análisis realizados 38

Ilustración 20: Relación de análisis realizados 39

Ilustración 21: Ejemplo de análisis estadístico en plataforma web del detector de malware 39

Ilustración 22: 20 mejores antivirus por número total de ficheros analizados 43

Ilustración 23: 20 peores antivirus por número total de ficheros analizados 44

Ilustración 24: 20 mejores antivirus por número total de malware detectado 47

Ilustración 25: 20 peores antivirus por número total de malware detectado 48

Ilustración 26: Clasificación de ficheros analizados por tipo de resultado (http://www.pcmag.com/article2/0,2817,2481367,00.asp) 49

Ilustración 27: Relación entre falsos positivos y falsos negativos. (Gestión de incidentes en seguridad informática. IFCT0109. Chicano Tejada, Esther. 2014).....50

Ilustración 28: Relación de falsos positivos por antivirus 53

Índice de tablas

Tabla 1: Definición tabla "*malware*" en la base de datos 18

Tabla 2: Definición tabla "*antivirus*" en la base de datos.....18

Tabla 3: Definición tabla "*machine*" en la base de datos 18

Tabla 4: Datos de ficheros analizados por antivirus con respecto al total (1000).....41

Tabla 5: Datos de ficheros considerados como positivos por antivirus 45

Tabla 6: Porcentaje de falsos negativos por antivirus.....51

Capítulo 1 Introducción

Bien es cierto que existen múltiples soluciones gratuitas y privativas, de naturaleza hardware (*Fortinet, PaloAlto, …*) y software (*Norton, Kaspersky, Panda, …*) que detectan malware. Usualmente, éstas se caracterizan por las siguientes peculiaridades:

- Son sistemas privativos que trabajan con una licencia por periodo de tiempo o licenciamiento permanente, pero tienen un coste asociado a la compra y renovación.
- Están encaminados a satisfacer las necesidades de una única plataforma. Es decir, las soluciones aplican a sistemas Microsoft Windows, Linux (u otros sistemas operativos), pero no a todos ellos en su conjunto.
- Requieren una instalación compilada expresamente para la plataforma a la que están diseñadas.
- El código fuente no es accesible ni para la lectura ni para la modificación.
- La controladora del antivirus debe estar en un servidor o máquina dedicada, dentro de la red local de la organización.
- No existe la posibilidad de detectar el malware en sistemas descentralizados.
- No proporcionan un histórico de cuántas veces un mismo software maligno ha sido detectado sobre una máquina determinada.
- Tampoco ofrecen la posibilidad de conocer cuántas máquinas diferentes han sido infectadas por un mismo malware.
- No aprovechan el conocimiento existente de diferentes fabricantes, técnicas y metodologías.
- Causan una gran ralentización de los sistemas, puesto que en la mayor parte de las ocasiones el análisis se realiza sobre la misma máquina en estudio.

Como se puede observar, existen limitaciones de índole técnico y organizativo que afectan a la eficacia y eficiencia de los sistemas de la información y equipos de seguridad.

Cabe destacar que un único antivirus no es capaz de detectar el 100% de los archivos maliciosos que se encuentran sobre un equipo[1]. Esto nos llevaría a pensar que si incrementamos el número de antivirus instalados mejoraríamos la eficiencia de las detecciones. Esto resulta poco aconsejable, pues:

- Los productos instalados consumirían grandes recursos: tiempos procesamiento, memoria, etc.

- Intentarían acceder a los mismos procesos y zonas de memoria para realizar acciones de control y monitorización sobre el sistema protegido. Esto podría causar bloqueos entre las aplicaciones.
- Un antivirus podría detectar las acciones de otro como malignas, dando lugar a falsos positivos y pérdida de funcionalidad de los mismos.

Adicionalmente, siempre se ha pensado que la seguridad sobre un sistema de la información depende única y exclusivamente de las primeras capas del protocolo de comunicaciones, sobre todo las referidas a las capas de red y transporte.

Sin embargo, buena parte de los problemas de seguridad se deben a la explotación de vulnerabilidades hardware o software mediante ficheros maliciosos. Son éstos a los que hay que prestar su debida atención para disponer de sistemas más confiables y seguros.

En el presente trabajo se muestra una solución que supera las barreras anteriormente descritas caracterizándose por su simplicidad y flexibilidad.

Capítulo 2 Objetivos

2.1 Objetivo general

La finalidad esencial del presente Trabajo Fin de Máster es proporcionar una técnica y herramienta de descubrimiento de malware que cumpla los siguientes objetivos:

- Proporcionar versatilidad en entornos multiplataforma; es decir, universalidad en su ejecución.
- Ofrecer respuesta en cualquier sistema de la información de una determinada organización independientemente de su naturaleza (público, privada, …) y ubicación.
- Aprovechar la fuente de conocimiento que proporcionan los diferentes fabricantes a través de la plataforma *VirusTotal* [i].
- Disminuir al máximo la ralentización que afecta a la eficacia y eficiencia de los dispositivos, así como a los procesos para los que está destinada. Para ello:
    - Realizar el análisis en ubicación ajena a la máquina analizada.
    - Utilizando técnicas y metodologías ligeras que permitan identificar de forma inequívoca el malware; como el uso de funciones hash criptográficas.
- Proporcionar una herramienta libre y gratuita.

Es necesario hacer hincapié en que la herramienta desarrollada permite detectar malware en entornos multiplataforma ya que su funcionalidad no se pierde si es ejecutado en diferentes sistemas operativos. Puede ser ejecutada en máquinas y servidores tanto Windows, Linux, como iOS.

De igual forma, buena parte de su potencialidad se debe a que los resultados quedan centralizados en una única base de datos.

Independientemente de la ubicación de los dispositivos que ejecuten el software, éstos reportarán los resultados a una máquina central que actuará a modo de servidor de base de datos.

Tal como se observa en el siguiente gráfico, máquinas de la oficina central (y satélites), fábricas, incluso dispositivos de empleados que estén realizando teletrabajo, podrán ser analizados remotamente y los resultados quedarán almacenados en el servidor de base de datos de la organización.

[i] https://www.virustotal.com/es/

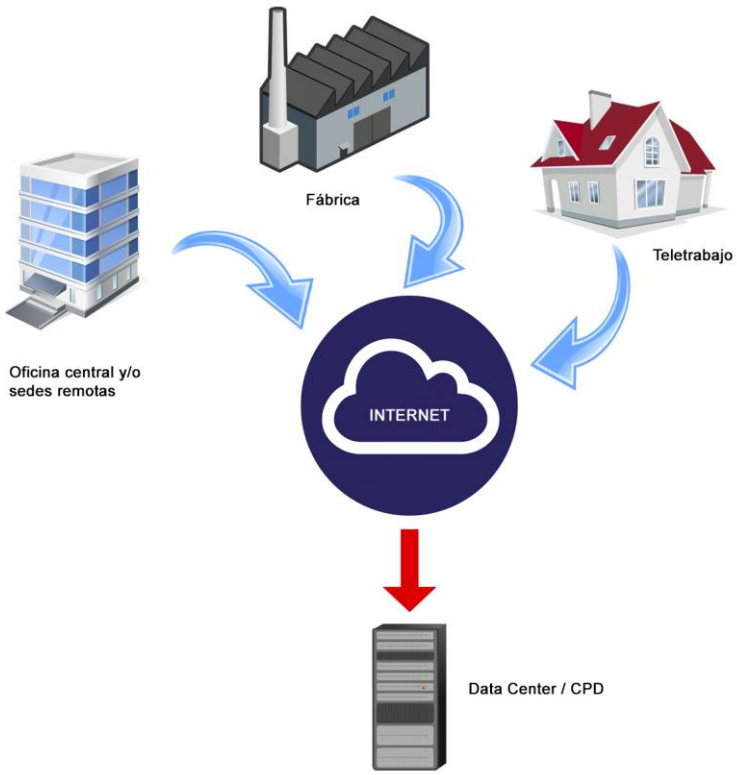


Ilustración 1: Estructura general

2.2 Objetivos específicos

Adicionalmente a los objetivos adicionales, con los resultados obtenidos se pretenderá:

- Evaluar la capacidad de análisis y descubrimiento de malware mediante los escaneos de cada antivirus.
- Proporcionar flexibilidad y suficiencia para personalizar la solución a cada organización o con ampliación de nuevas funcionalidades.

Capítulo 3 Metodología

3.1 Firmas y funciones hash criptográficas

Una firma es un patrón escogido de un virus que ha sido creado a raíz del análisis del malware mediante un escáner. La firma suele contener el código de solución o la propia vacuna contra el virus. Es por ello que existen ficheros infectados que pueden ser reparados a un estado estable anterior.

La totalidad de firmas conforman una base de datos que el antivirus utiliza para identificar ficheros maliciosos. Cada antivirus dispone de su propia relación de firmas.

Actualmente, también se tienen en cuenta otros métodos de detección de malware a parte de las firmas, como las técnicas heurísticas que permiten detectar variaciones de virus conocidos. Pero las firmas siguen siendo la técnica más funcional de todos los antivirus.

Para el detector de malware utilizaremos una cadena alfanumérica asociada a cada fichero que, a su vez, nos servirá de identificador inequívoco.

Esta cadena será generada mediante la función hash criptográfica SHA-256.

Las funciones criptográficas (también denominadas funciones “hash”)[2] son un algoritmo matemático que transforma cualquier bloque arbitrario de datos de tamaño N en una nueva cadena de caracteres de longitud fija de tamaño M. Esta longitud dependerá del algoritmo o función hash empleada.

No existe límite para N, pero M siempre tendrá el mismo tamaño para un mismo algoritmo usado. Es lo que se conoce con el concepto de “compresión”.

Una de las características de este tipo de algoritmos es que son unidireccionales[3]. Es decir, a través de la cadena resultante M es imposible conocer cuáles son los datos N originales.

Las funciones hash disponen, además, de ciertas propiedades[4] que posibilitan el desarrollo de la herramienta de detección de malware. A continuación quedan indicadas, ya adaptadas al propósito de este trabajo:

- Facilidad de cálculo: Es muy fácil calcular M a partir de N.

- Difusión e integridad: M debe ser calculado a través de una función compleja sobre N (hash(N)). Si se modifica un único bit de los datos originales N, bebería cambiar aproximadamente la mitad de los bits de M.
- Deterministas: Un mismo fichero malicioso siempre proporciona el mismo valor hash.
- Colisión débil: Es computacionalmente imposible encontrar un par de ficheros maliciosos N y N´ de forma que al aplicar las funciones hash sobre cada una de ellos se obtenga el mismo resumen M. Es decir, encontrar hash(N) = hash(N´) es complejo.

El detector de malware desarrollado hace uso de las función hash criptográfica SHA-256 (de la familia SHA, publicada por el *Instituto Nacional de Estándares y Tecnología - NIST*[ii]) para identificar de forma única e inequívoca a cada fichero analizado.

La función hash SHA-256 aplicada sobre una entrada N, proporciona un resultado M de longitud de 256 bits, que normalmente son expresados como 64 dígitos en caracteres hexadecimales.

Por ejemplo, suponiendo un fichero que está formado por la siguiente sentencia maliciosa:

```
Esto es el código de un virus
```

Obtenemos el valor mostrado a continuación aplicándole la función SHA-256:

```
3a9f1431fa7f53ba96b19b524b81eb149e47f34df7c59e9614160d25f53d54ce
```

La utilidad de las funciones criptográficas es proporcionar la funcionalidad básica y esencial del software detector de malware.

Como se ha comentado anteriormente, permiten calcular de forma rápida y eficaz un valor único para una determinada entrada. Por tanto, pueden asemejarse a una firma que identifican a cada uno de los ficheros analizados.

Esta firma es usada para consultar a una base de datos local o al repositorio del proyecto *VirusTotal* si tiene constancia de elementos maliciosos sobre un determinado fichero.

[ii] http://www.nist.gov/

Por tanto, la idea parece clara: es más ágil y efectivo consultar por una cadena de 256 bits que por un fichero malicioso que desconocemos el tamaño pero que pueden superar varios megabytes.

Aplicándolo al ejemplo anterior, se preguntará a un repositorio de malware si el fichero con hash "`3a9f1431fa7f53ba96b19b524b81eb149e47f34df7c59e9614160d25f53d54ce`" es considerado malicioso.

3.2 Proyecto VirusTotal

VirusTotal es una empresa de Málaga (España) que, en el año 2012, fue comprada por el gigante americano *Google*[5].

VirusTotal proporciona una herramienta sencilla para detectar malware en ficheros y páginas web basado en la experiencia y reconocimiento de más de sesenta antivirus.

Su principal característica es que se trata de un servicio gratuito para todos los internautas que desean analizar un archivo[6].

La subida de fichero para su análisis se puede efectuar mediante tres vías:

1. Manual: a través de la plataforma web
2. Análisis mediante URL
3. Mediante API

Para el desarrollo del detector de malware se ha utilizado una API desarrollada en Perl.

Se trata de un lenguaje de programación interpretado (no compilado), software libre y está licenciado bajo la *GNU Public License*. Existen distribuciones para la mayor parte de sistemas operativos y especialmente extendido en Unix.

La API proporciona mayor versatilidad, puesto que permite la consulta de los resultados fruto de análisis de ficheros previamente analizados, así como la subida de archivos en caso que no exista un análisis anterior en el repositorio de *VirusTotal*.

La consulta se realiza mediante una función hash criptográfica, que como hemos comentado en el punto anterior, se utilizará SHA-256. Bien es cierto que *VirusTotal* permite la consulta por varias funciones resumen, pero he estimado usar SHA-256 frente a MD5 o SHA-1,

puesto que los dos últimos generan una cadena resumen de 128 bits y 160 bits, respectivamente, frente a los 256 de SHA-256. Esto significa que la probabilidad de colisión (es decir, dos ficheros diferentes que generen la misma función resumen) es mucho menor en SHA-256 que MD5 o SHA-1.

Por tanto, la potencia de usar la API radica en poder realizar consultas de ficheros de forma remota mediante una función hash.

Así mismo, si no existe referencia al fichero a analizar en las bases de datos del proyecto *VirusTotal*, éste podrá ser subido a la plataforma mediante el protocolo HTTP-POST.

El desarrollo de la API está inspirada en un script[iii] del PhD. Christopher Frenz y la especialista en seguridad Michelle Sullivan.

3.3 Estructura

El detector de malware consta de los siguientes elementos:

1. Cliente (detector)
2. Sistema gestor de base de datos MySQL (SGBD MySQL)
3. Servidor web (opcional)

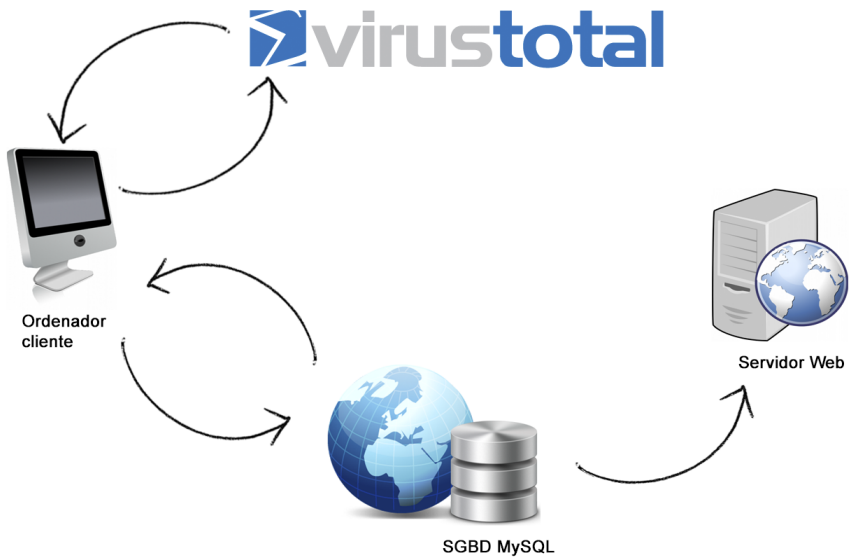


Ilustración 2: Estructura particular

[iii] http://perlgems.blogspot.com.es/2012/05/using-virustotal-api-v20.html

3.3.1 Cliente

El cliente se compone de tres ficheros claramente diferenciados, todos ellos desarrollados en Perl.

- *search.pl*: Función principal del programa. Invoca a funciones de los ficheros "*librería.pm*" y "*VTMySQL.pm*".
  Permite analizar un archivo específico o todos los contenidos en un directorio.

- *VirusTotal.pm*: API que interactúa con el proyecto *VirusTotal*. En ella está incluida la licencia proporcionada para subir archivos a la plataforma.

- *VTMySQL.pm*: API que proporciona la comunicación con una base de datos de MySQL. En ella están incluidos los parámetros de conexión a la base de datos, definidos como variables al inicio de la librería.

3.3.2 Sistema gestor de base de datos MySQL

El sistema gestor de base de datos MySQL (SGBD MySQL) almacena tres tablas que contienen la información obtenida proporcionada mediante los análisis de *VirusTotal*. Esta base de datos es local e interna a la organización, ajena al proyecto de *Google*.

El motivo de disponer una base de datos local, adicionalmente al repositorio proporcionado por el proyecto *VirusTotal,* se debe a cuatro motivos esenciales:

- Las consultas a *VirusTotal* están limitadas a un número máximo de solicitudes por día, desde una misma dirección IP o licencia.
- Realizar las consultas a *VirusTotal* conllevarán más tiempo que aquellas realizadas a una base de datos interna de la propia organización.
- Las consultas a *VirusTotal* pueden producir problemas de time-out en las respuestas, así como indisponibilidad en el servicio del proyecto. Ambos hándicap quedan solventados con bases de datos locales debidamente redundadas.
- Finalmente, se logra disponer de un histórico y trazabilidad manteniendo información de qué ficheros contienen malware y qué máquinas han sido afectadas, incluyendo los tiempo de detección y antivirus que participaron en el análisis.

Tabla 1: Definición tabla "*malware*" en la base de datos

malware		
Columna	**Tipo de dato**	**Descripción**
id	int(10)	Clave primaria. Número autoincremental
genericName	varchar(100)	Nombre genérico del malware
sha256	char(64)	Función hash criptográfica SHA256 aplicada sobre el fichero analizado
positives	int(10)	Número de antivirus cuyo resultado del análisis ha sido positivo
total	int(10)	Número de antivirus que han analizado el fichero

Tabla 2: Definición tabla "*antivirus*" en la base de datos

antivirus		
Columna	**Tipo de dato**	**Descripción**
id	int(10)	Clave primaria. Número autoincremental
antivirus	varchar(100)	Nombre comercial del antivirus
positives	int(11)	Número de positivos con respecto al número total de análisis efectuados
total	int(11)	Número total de análisis efectuados

Tabla 3: Definición tabla "*machine*" en la base de datos

machine		
Columna	**Tipo de dato**	**Descripción**
id	int(11)	Clave primaria. Número autoincremental
malware_id	int(11)	Clave extranjera. Refiere a la clave primaria del malware
name	varchar(50)	Nombre de la máquina sobre la que se efectúa el análisis
path	varchar(100)	Ruta absoluta donde se encuentra el malware
detection	timestamp	Fecha y hora de detección del malware

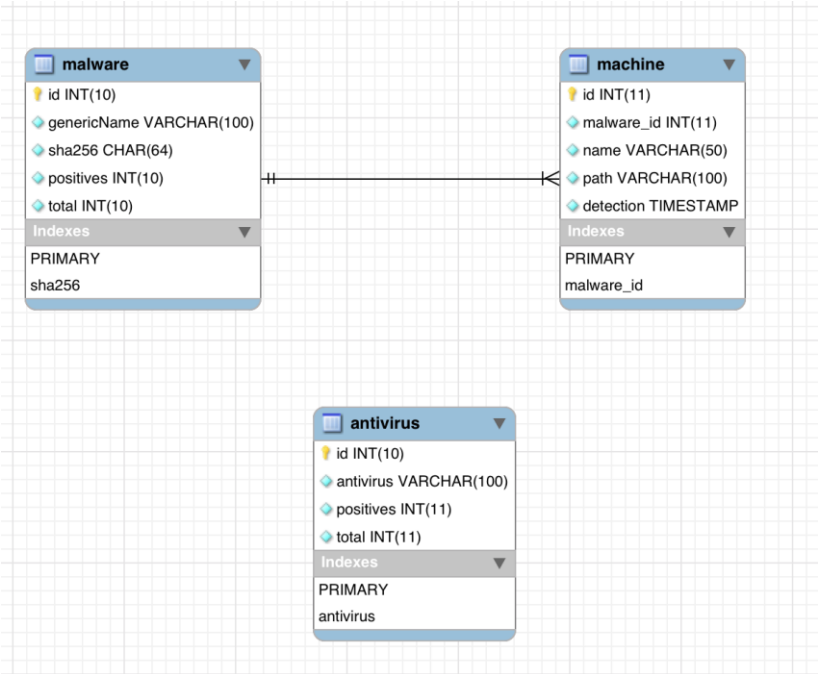


Ilustración 3: Diagrama entidad-relación

3.3.3 Servidor web

Completamente opcional al la estructura del detector de malware, se permite incluir a la arquitectura un servidor web que interactúe con los datos almacenados en la base de datos.

Se aconseja su inserción, pues a parte de permitir visualizar los datos de forma más clara, concisa y visualmente más agradable, proporciona la capacidad para disponer la información a usuarios, separados por roles.

Como elemento más destacable es que no requiere que un usuario interactúe directamente con la base de datos, evitando así, cualquier error fortuito o circunstancial en el manejo de la información.

La herramienta detector de malware proporciona una interfaz sencilla pero útil para visualizar la información más destacable sobre ficheros maliciosos de una organización.

Capítulo 4 Flujo de datos

4.1 JSON

JSON[7] (*JavaScript Object Notation*) es un formato ligero de intercambio de datos. Su compresión es muy sencilla por parte de humanos, así como su interpretación o generación mediante máquinas.

Es un formato de texto independiente del lenguaje, pero utiliza convenciones ampliamente conocidos por los desarrolladores de la familia de lenguajes C, JavaScript, Perl, Python, …

JSON dispone de dos estructuras. La primera de ella es una colección de nombres/valor y, en segundo lugar una lista ordenada de valores (en otros lenguajes se refiere a los arrays, vectores, listas o secuencias).

4.2 EICAR

EICAR es un archivo de prueba antimalware que ha sido desarrollado por el Instituto Europeo para la Investigación de los Antivirus Informáticos (EICAR) [8].

Se trata de un pequeño archivo COM de 68 bytes que los sistemas antivirus lo detectan como malware a pesar que no lo es.

Este fichero permite verificar que el antivirus está funcionando correctamente, ya que los está diseñado para que reaccionen a él como si de un virus real se tratase. Es catalogado con el nombre "*EICAR test file*" [9].

El archivo COM debe contener la siguiente línea:

```
X5O!P%@AP[4\PZX54(P^)7CC)7}$EICAR-STANDARD-ANTIVIRUS-TEST-FILE!$H+H*
```

EICAR es utilizado en el detector de malware para determinar que la API (*VirusTotal.pm*) y la plataforma del proyecto de Google están funcionando correctamente, tanto las conexiones como el proceso de análisis.

4.3 Flujo general

La idea original del detector de malware es bastante sencilla y eficaz, lo que facilita su desarrollo.

1. En primera instancia se lee el fichero y, conforme a ello, se calcula su hash SHA-256.
2. Con el resultado de la función hash (en adelante, *resumen*) se consulta a la base de datos local (propia de la organización) si ya se encuentra registrado el fichero. En caso afirmativo se registrará de nuevo la detección de malware sobre la máquina afectada. En otro caso, se lanzará el escáner del fichero contra *VirusTotal*.
3. Invocada la llamada al método público "*scan()*", si es la primera vez que se realiza la conexión a *VirusTotal* desde el inicio de la ejecución del detector de malware o si han pasado más de 300 segundos (5 minutos) desde la última conexión, se realiza una prueba de conectividad y funcionamiento de la plataforma con el hash del archivo de prueba antimalware *EICAR*.
4. Si la conexión es correcta, seguidamente se realiza la subida del fichero a la plataforma *VirusTotal* mediante la API 2.0 desarrollada en Perl. Una vez subido, se espera el resultado.
5. Si el resultado existe en el mismo momento, éste será devuelto en un objeto JSON y la información que él contiene se agregará

   Si el resultado tiene una demora, fruto del tiempo que toman los antivirus para su análisis, será consultado el reporte final en franjas de 180 segundos por fichero hasta que el objeto JSON con todos la información sea devuelto por *VirusTotal*. También, en este caso, los datos serán agregados a la base de datos local para futuras consultas.

4.3.1 Flujograma

El diagrama de flujo se muestra a continuación, que corresponde con el flujo general explicado en el anterior punto:

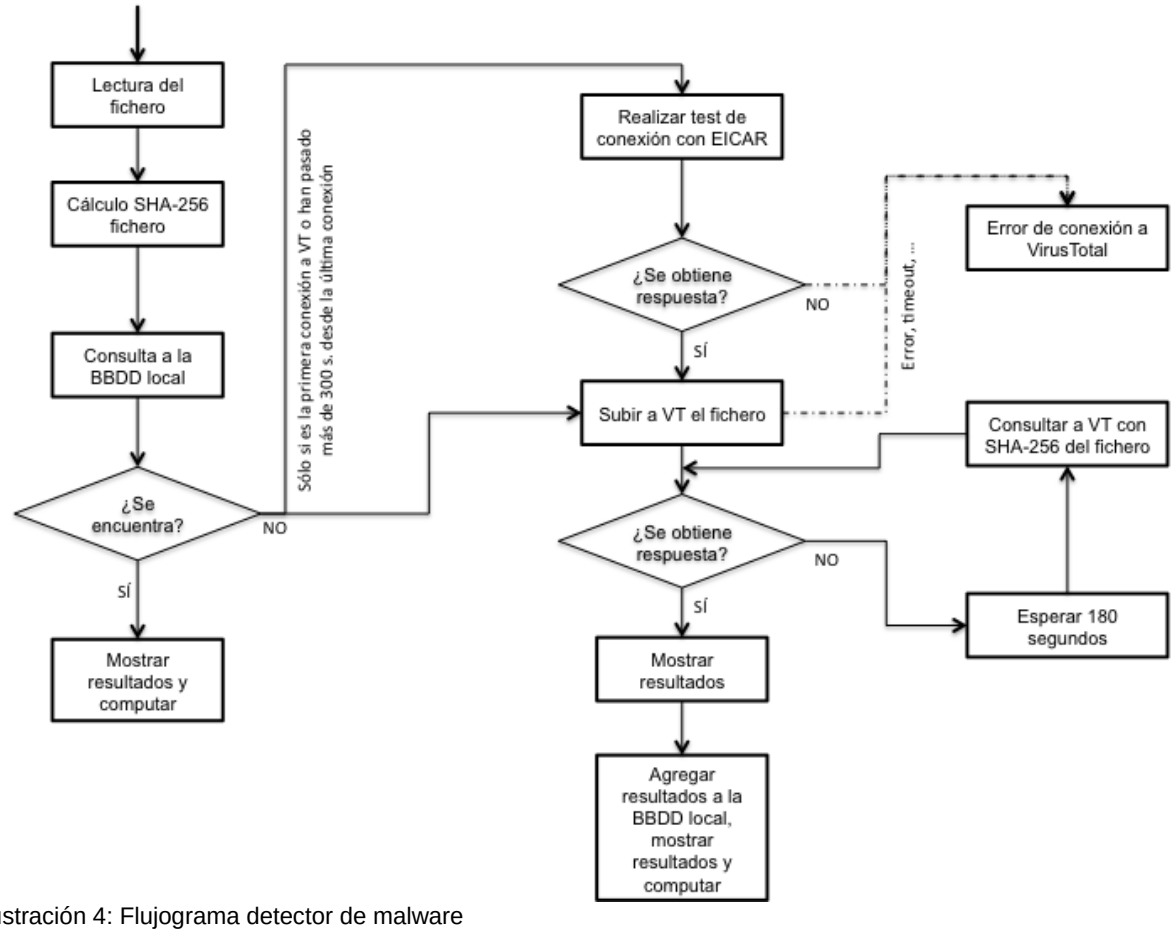


Ilustración 4: Flujograma detector de malware

Como se puede observar en el flujograma, la primera vez siempre se realiza en análisis del fichero mediante su contenido. Esto es debido a que se desconoce, a ciencia cierta, si existe una entrada asociada en *VirusTotal* o cuándo fue la última vez en la que se realizó su análisis (pudiendo haber sido hace varios meses o años).

Para ello, actualizamos el reporte que pueda proporcionar un nuevo análisis de cada antivirus.

En el caso que el análisis se demore por varios minutos, éstos serán de nuevo solicitados en franjas de 180 segundos (3 minutos) mediante, ahora sí, la función criptográfica SHA-256 del fichero analizado.

Capítulo 5 Ejecución y resultados

5.1 Requisitos

Aunque el detector de malware es considerado multiplataforma, en el presente trabajo se hizo mención especial a la implementación y funcionamiento sobre máquinas Linux.

Por tanto, los requisitos para poder ejecutar el software, son:

- SGBD MySQL que permita conexiones desde el exterior
    - Usuario y contraseña con privilegios para la ejecución de comandos "*SELECT*", "*UPDATE*" e "*INSERT*". Estas credenciales serán usadas en el programa detector de malware.
    - Usuario y contraseña con privilegios para la ejecución de comandos "*SELECT*". Estas credenciales serán usadas en el servidor web para la visualización de datos. Por motivos de seguridad, se aconseja aplicar el principio de mínimo privilegio y separación de funcionalidades; es por ello, por lo que no se utiliza el mismo usuario que el usado para el detector de malware.
- Máquina cliente donde se ejecutarán el detector de malware. Los test finales han sido realizados sobre una máquina con sistema operativo *CentOS release 6.6 (Final)*.

Paquetes instalados con CPAN `(perldoc perllocal)`	**Paquetes instalados con RPM** `(rpm -qa | grep -i perl)`
Algorithm::Diff, 1.1903	perl-version-0.77-136.el6_6.1.i686
Test::Simple, 1.001014	perl-Pod-Simple-3.13-136.el6_6.1.i686
Parent, 0.232	perl-5.10.1-136.el6_6.1.i686
URI, 1.67	perl-Digest-SHA-5.47-136.el6_6.1.i686
LWP::MediaTypes, 6.02	perl-ExtUtils-MakeMaker-6.55-136.el6_6.1.i686
Encode::Locale, 1.04	perl-devel-5.10.1-136.el6_6.1.i686
IO::HTML, 1.001	perl-IO-Compress-Base-2.021-136.el6_6.1.i686
HTTP::Date, 6.02	perl-IO-Compress-Zlib-2.021-136.el6_6.1.i686
File::Listing, 6.04	perl-HTML-Tagset-3.20-4.el6.noarch
HTML::Tagset, 3.20	perl-libwww-perl-5.833-2.el6.noarch
WWW::RobotRules, 6.02	perl-DBI-1.609-4.el6.i686

Spiffy, 0.46	perl-JSON-2.15-5.el6.noarch
Text::Diff, 1.41	perl-Pod-Escapes-1.04-136.el6_6.1.i686
Test::Base, 0.88	perl-libs-5.10.1-136.el6_6.1.i686
Test::YAML, 1.05	perl-Module-Pluggable-3.90-136.el6_6.1.i686
YAML, 1.14	perl-Test-Harness-3.17-136.el6_6.1.i686
ExtUtils::MakeMaker, 7.04	perl-ExtUtils-ParseXS-2.2003.0-136.el6_6.1.i686
Perl::OSType, 1.008	perl-CPAN-1.9402-136.el6_6.1.i686
Locale::Maketext::Simple, 0.21	perl-Compress-Raw-Zlib-2.021-136.el6_6.1.i686
Params::Check, 0.38	perl-Compress-Zlib-2.021-136.el6_6.1.i686
Versión, 0.9912	perl-URI-1.40-2.el6.noarch
Module::Metadata, 1.000026	perl-HTML-Parser-3.64-2.el6.i686
Module::Load, 0.32	perl-Time-HiRes-1.9721-136.el6_6.1.i686
Module::CoreList, 5.20150320	perl-DBD-MySQL-4.013-3.el6.i686
Module::Load::Conditional, 0.64	perl-Crypt-SSLeay-0.57-17.el6.i686
IPC::Cmd, 0.92	
ExtUtils::CBuilder, 0.280220	
Net, 3.06	
JSON, 2.90	
Try::Tiny, 0.22	
CPAN::Meta::YAML, 0.014	
Parse::CPAN::Meta, 1.4414	
CPAN::Meta, 2.150001	
Mozilla::CA, 20141217	
HTTP::Negotiate, 6.01	
HTTP::Daemon, 6.01	
Net::http, 6.07	
HTTP::Cookies, 6.01	
LWP, 6.13	

- Número de licencia: *VirusTotal* proporciona las licencias de forma gratuita, pero el usuario u organización debe estar previamente registrado en el proyecto e indicar los motivos de uso mediante una solicitud formal (*https://www.virustotal.com/* → *"Únete a la comunidad*").

  En caso de no disponer de licencia, existe un límite técnico de 4 peticiones por minuto.

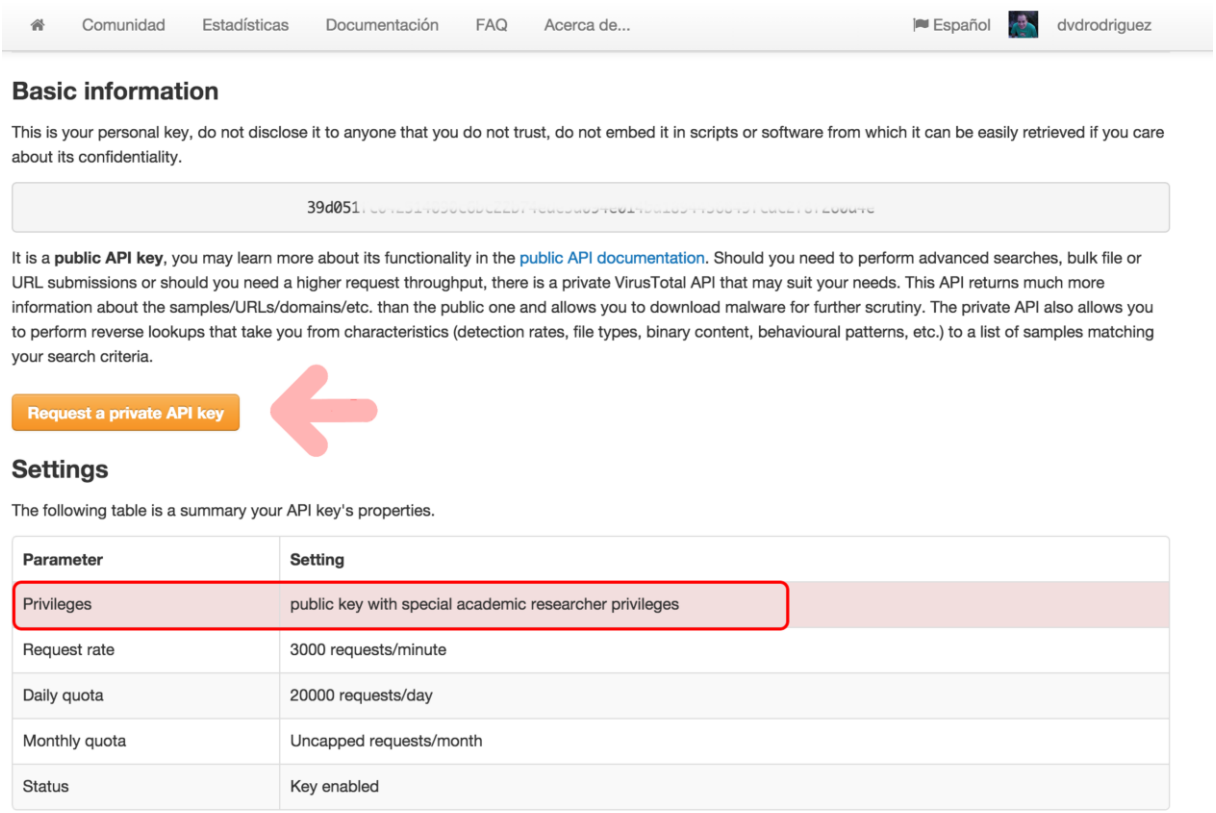


Ilustración 5: Área privada y solicitud de licencia a VirusTotal

- Opcionalmente, se aconseja un servidor web en el que mostrar los datos de forma más interactiva que la ejecución de *queries* directamente sobre el SGBD. Este servidor web deberá tener acceso a la base de datos para realizar consultas del tipo "*SELECT*".

5.2 Parametrización e invocación

Como se ha comentado anteriormente, el software detector de malware está compuesto de tres ficheros que deben ser agregados en cada cliente.

La peso de todos ellos en su conjunto no supera los 50 KB, por lo que se trata de un cliente extremadamente ligero. Al estar desarrollado en Perl, lenguaje interpretado, no requiere compilación previa.

La configuración básica a realizar se encuentra en el fichero denominado "*VTMySQL.pm*". En él es necesario completar las siguientes variables:

1. `$_MACHINE_NAME_`: nombre completo de la máquina. Se aconseja que se el nombre se encuentre en formado FQDN (nombre del equipo + dominio asociado al equipo), sobre todo si la organización dispone de múltiples dominios.
   De esta forma, será muy fácil poder identificar la máquina.
2. `$dbhost`: Dirección del servidor MySQL. Bien puede tratarse de una dirección con resolución DNS directa o propiamente la dirección IP del mismo.
   Si el puerto de conexión fuera diferente al estándar (TCP/3306), la dirección del servidor deberá ser del tipo *"*`$dbhost:port`". Por ejemplo: *hosting02.servidor.com:3309*
3. `$dbname`: Nombre de la base de datos o esquema de la misma.
4. `$dbuser`: Usuario utilizado para la autenticación.
5. `$dbpwd`: Contraseña del usuario anterior.

Ilustración 6: Ejemplo de configuración de parámetros de conexión (*VTMySQL.pm*)

La ubicación de los ficheros puede ser donde el responsable del sistema lo desee, aunque para máquinas Linux se aconseja en el directorio `/usr/local/bin`, siguiendo el estándar de instalación de software local.

Observación: Si las librería ("*VirusTotal.pm"* y "*VTMySQL.pm"*) se encuentran en un directorio ajeno al programa principal ("*search.pl"*), esto deberá ser indicado en el fichero "*search.pl*", en la sentencia "`use lib 'path'`", siendo 'path' la ruta absoluta donde se encuentren dichas librerías.

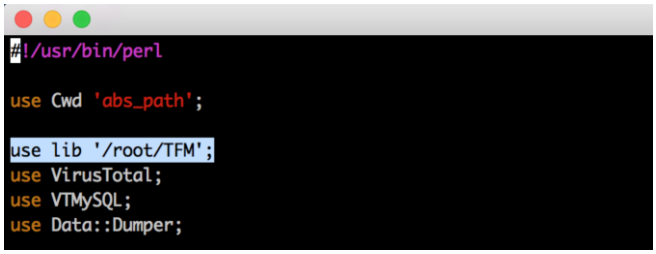


Ilustración 7: Cambio de ubicación de las librerías del detector de malware (*search.pl*)

Una vez configurados los parámetros de conexión, la ejecución del software es tremendamente sencilla.

Se pueden comprobar dos modalidades de ejecución:

- Análisis de un único fichero: La ejecución se realizará invocando al programa principal con el flag "`-f`" seguido del nombre del fichero incluyendo su ruta absoluta (absolute_path/file_to_analyse).

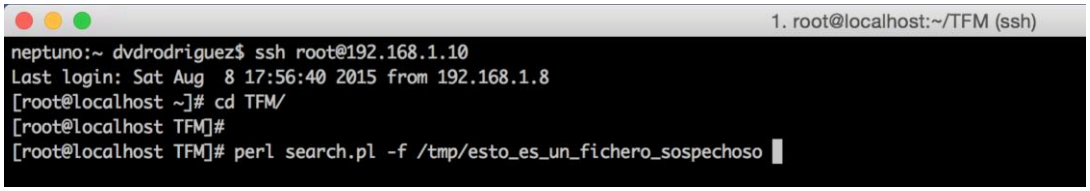


Ilustración 8: Ejemplo de ejecución de análisis de un fichero

- Análisis de un directorio completo: En este caso la ejecución se realizará invocando al programa principal con el flag "`-d`" seguido de la ruta absoluta al directorio (absolute_path_of_a_directory).
  Cabe destacar que se excluyen del directorio a analizar los elementos "*.*" y "*..*" de los sistemas Unix, así como "*.DS_Store*", propio de sistemas iOS.

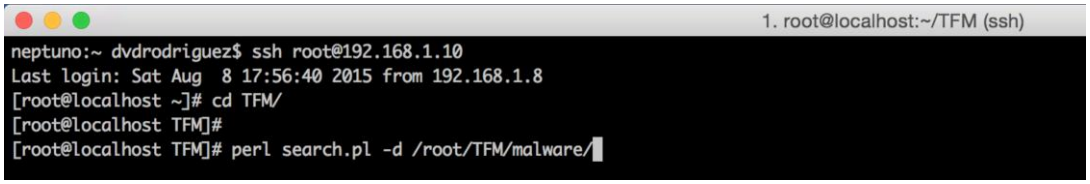


Ilustración 9: Ejemplo de ejecución de análisis de un directorio

El resultado –o sentencia- definitiva sobre si el fichero analizado se considera malware o no, es mostrado por la salida estándar de la máquina donde se ejecuta el programa detector de malware.

En primera instancia, en el ejemplo que se muestra a continuación, se ejecuta el análisis de un fichero que no se encuentra en la base de datos local ni en *VirusTotal*. Por tanto, *VirusTotal* realizará en análisis del fichero y dispondrá un veredicto definitivo en el que se indique si éste se considera malicioso, o no.

El usuario percibirá que el tiempo de cómputo es considerablemente alto, pues después de realizar la consulta a la base de datos local, se realizará la petición a *VirusTotal* subiendo el contenido del mismo a la plataforma y realizándose un análisis completo del fichero.

Seguido al análisis se dispondrá de los resultados proporcionados por cada uno de los antivirus. Si estos reportes se demoran en demasía serán solicitados concurrentemente en periodos de 3 minutos.

En la línea donde se muestra el resultado definitivo del análisis, se indica la entidad que ha proporcionado la información. En este caso, se puede comprobar que el detector de malware indica `[VirusTotal]`. Es correcto, pues como hemos comentado se trata del primer análisis realizado.

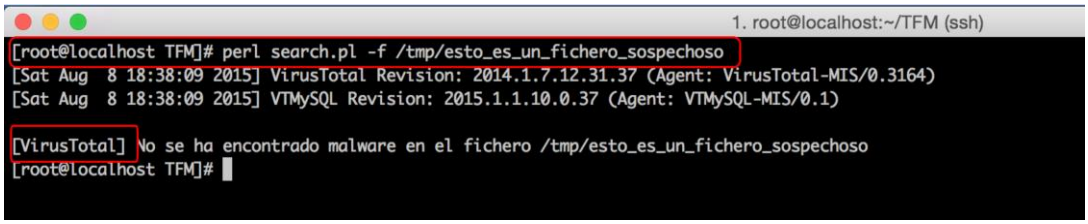


Ilustración 10: Análisis de un fichero no tratado previamente

Si volvemos a solicitar el análisis del mismo fichero, podremos comprobar que el resultado se muestra de forma inmediata. Como hubo un análisis previo que se almacenó en la base de datos local, la respuesta es proporcionada por el SGBD de la propia organización. Es por ello que, entre corchetes, se muestra el indicador `[local]`.

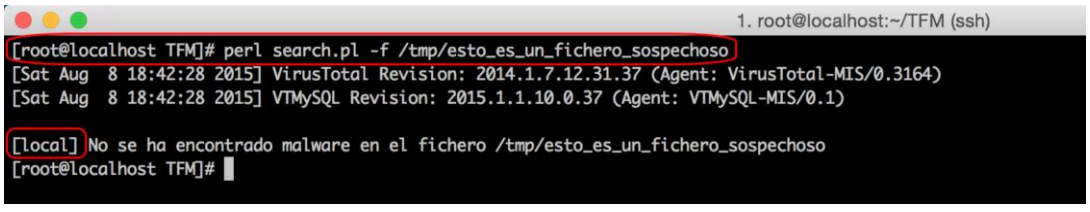


Ilustración 11: Análisis de un fichero tratado previamente

5.3 Resultados

Las respuestas proporcionadas por *VirusTotal*, independientemente que la consulta se haya realizado mediante función hash criptográfica SHA-256 o a través de la subida del propio fichero, son devueltas con un objeto en lenguaje JSON.

Las respuestas que recibe el detector de malware una vez se ha realizado la consulta mediante HTTP-POST pueden ser de tres tipos:

- Código de respuesta -2: Se ha realizado una petición de análisis, pero ésta se encuentra encolada.
- Código de respuesta 0: El fichero por el que se hace la consulta no se encuentra aún en el sistema.

```
Enviando hash (02493311e5b3cc658a1d7f38b2dbcb05ab270beb9943351a9cd0aa360198b4bc).. at
/root/TFM/VirusTotal.pm line 565.
Respuesta obtenida:
{
    "response_code": 0,
    "resource": "02493311e5b3cc658a1d7f38b2dbcb05ab270beb9943351a9cd0aa360198b4bc",
    "verbose_msg": "The requested resource is not among the finished, queued or pending scans"
}
```

Si en el momento en el que está encolado solicitamos el reporte a través de la aplicación web, obtendremos el mismo mensaje que el devuelto en el objeto JSON.

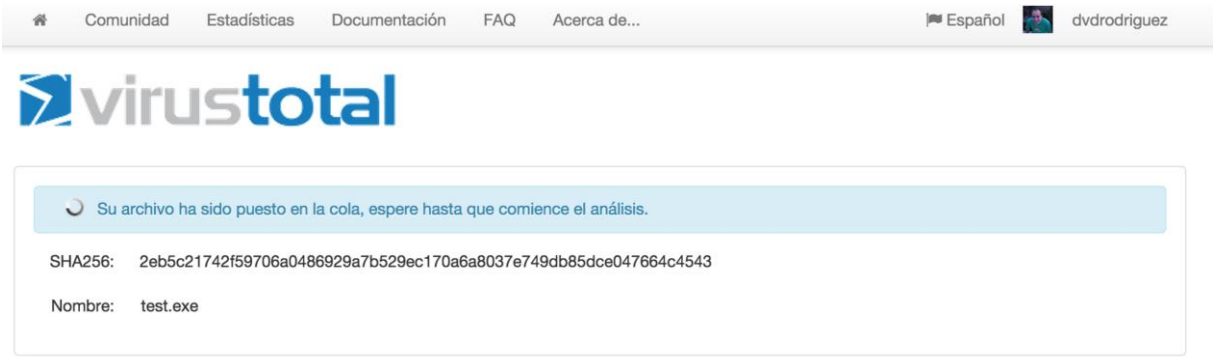


Ilustración 12: Petición encolada en VirusTotal

- Código de respuesta 1: El fichero se ha subido con éxito, ha sido analizado y la respuesta se encuentra disponible. Información completa referida al análisis del fichero.

```
Enviando hash (02493311e5b3cc658a1d7f38b2dbcb05ab270beb9943351a9cd0aa360198b4bc).. at
/root/TFM/VirusTotal.pm line 565.
Respuesta obtenida:
{
"scans":
    {
        "Bkav": {"detected": false, "version": "1.3.0.7062", "result": null, "update": "20150812"},
        "CMC": {"detected": false, "version": "1.1.0.977", "result": null, "update": "20150710"},
        ...
    },
    ...
    "response_code": 1, "scan_date": "2015-08-12 12:04:35",
    "permalink":
"https://www.virustotal.com/file/02493311e5b3cc658a1d7f38b2dbcb05ab270beb9943351a9cd0aa360198b4bc/a
nalysis/1439381075/",
    "verbose_msg": "Scan finished, information embedded",
    ...
}
```

- Otros códigos de respuesta: Mensajes de error en caso existan inconvenientes debido a la velocidad de solicitud (estado http 204) o llamadas a funciones en las que no se tienen privilegios (estado http 403), como el exceso en el número de solicitudes por minuto o haber sobrepasado los límites aceptados por la licencia.

A continuación se muestra en detalle la estructura de la respuesta satisfactoria obtenida de una petición de un fichero malicioso.

- [Color amarillo]: Ejemplo del análisis del antivirus "*Kaspersky*". En la respuesta se indica la versión, fecha de subida (y último análisis del fichero) y el nombre genérico del malware. En este caso, el fichero analizado es un gusano para Windows transmitido por correo electrónico.
- [Color verde]: SHA-256 del fichero subido a *VirusTotal*.

- [Color rojo]: Número total de antivirus que participan en el análisis y cuántos de ellos han proporcionado un positivo. En este caso, un total de 56 antivirus han realizado el análisis, pero sólo 50 lo detectan como malicioso.

```
$VAR1 = {
'scans' => {
                      'Microsoft' => {
                                       'version' => '1.1.11903.0',
                                       'update' => '20150808',
                                       'result' => 'Worm:Win32/Mydoom.O@mm',
                                       'detected' => $VAR1->{'scans'}{'Qihoo-360'}{'detected'}
                                     },
                      'Kaspersky' => {
                                       'version' => '15.0.1.10',
                                       'update' => '20150808',
                                       'result' => 'Email-Worm.Win32.Mydoom.m',
                                       'detected' => $VAR1->{'scans'}{'Qihoo-360'}{'detected'}
                                     },
                      'Avast' => {
                                   'version' => '8.0.1489.320',
                                   'update' => '20150808',
                                   'result' => 'Win32:Mydoom-M [Wrm]',
                                   'detected' => $VAR1->{'scans'}{'Qihoo-360'}{'detected'}
                                 },
                      'Ad-Aware' => {
                                      'version' => '12.0.163.0',
                                      'update' => '20150808',
                                      'result' => 'Worm.Generic.24461',
                                      'detected' => $VAR1->{'scans'}{'Qihoo-360'}{'detected'}
                                    },
                      'Panda' => {
                                   'version' => '4.6.4.2',
                                   'update' => '20150808',
                                   'result' => 'W32/Mydoom.N.worm',
                                   'detected' => $VAR1->{'scans'}{'Qihoo-360'}{'detected'}
                                 }

                       ... todos los antivirus que han realizado el análisis...
                    },
          'resource' => '0b7bbb808d47de269f829f110a0078c9ee27b09d021a531139b719fea2d0f4a4',
          'md5' => '05a3842e85af3e458f24270a351e407e',
          'sha1' => '62a56dc27f946745000a9fe9f6935fcc86088b3e',
          'sha256' => '0b7bbb808d47de269f829f110a0078c9ee27b09d021a531139b719fea2d0f4a4',
          'positives' => 50,
          'total' => 56,
          'scan_date' => '2015-08-08 11:29:02',
          'verbose_msg' => 'Scan finished, information embedded'
          'scan_id' => '0b7bbb808d47de269f829f110a0078c9ee27b09d021a531139b719fea2d0f4a4-1439033342',
          'response_code' => 1,
          'permalink'                                                                                    =>
'https://www.virustotal.com/file/0b7bbb808d47de269f829f110a0078c9ee27b09d021a531139b719fea2d0f4a4/analysis/1439033342
/',
        };
```

Una vez obtenido los resultados por parte de *VirusTotal*, éstos son almacenados en la base de datos local. De esta forma, la siguiente vez que se solicite el análisis del mismo fichero (misma función resumen SHA-256), no será necesario consumir una petición de las proporcionadas por el proyecto de Google en su licencia y, por tanto, el tiempo de cómputo será mucho menor.

Adicionalmente a los resultados proporcionados en lenguaje JSON, en el área personal del proyecto se pueden verificar qué ficheros han sido analizados.

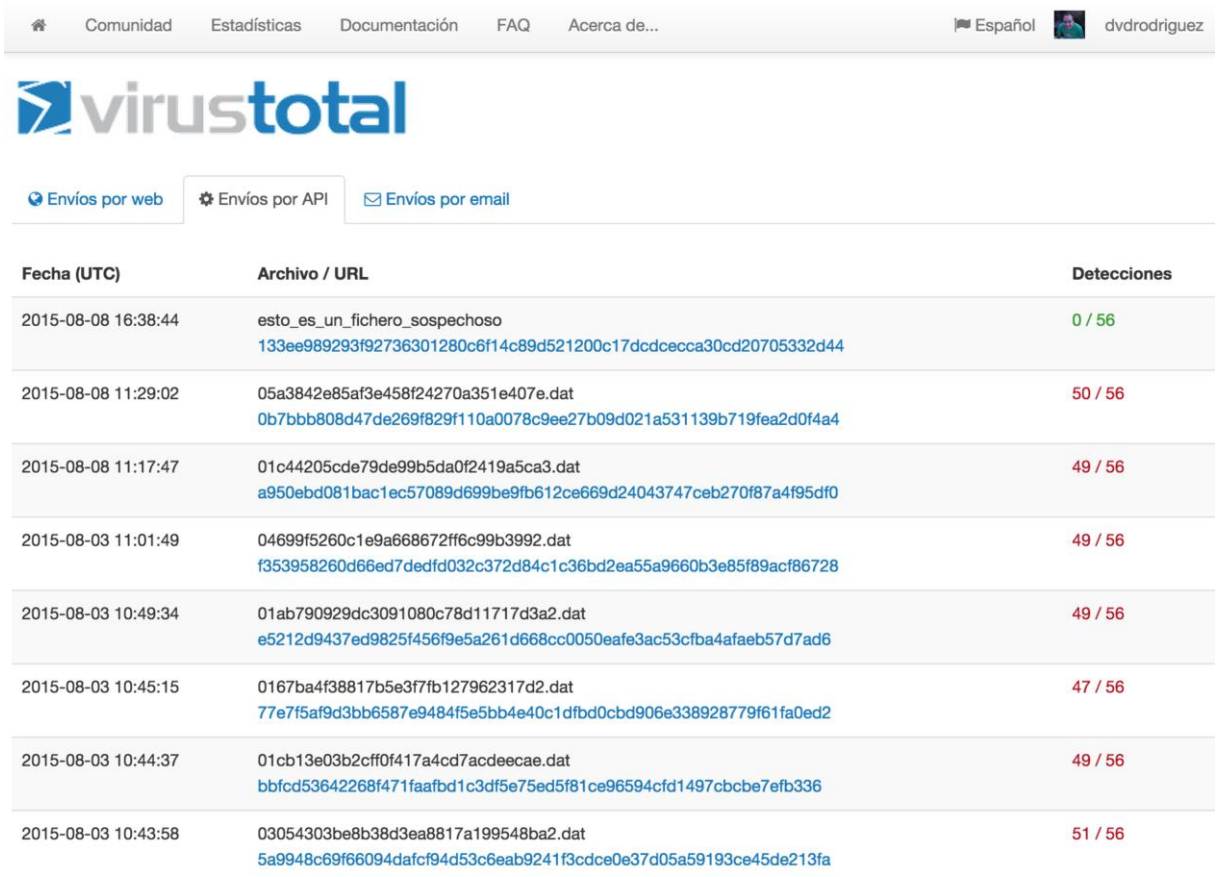


Ilustración 13: Relación de ficheros subidos a *VirusTotal* mediante la API

También resulta interesante poder visualizar las peticiones realizadas por día.

VirusTotal discierne sobre las consultas realizadas para el análisis y las únicamente referidas a la solicitud de reportes de ficheros (independientemente hayan sido escaneados previamente, o no).

De esta forma se mantiene controlado si se excede del número de solicitudes autorizadas por la licencia, que no deberá exceder de la suma de ambos contadores. En otro caso, se producirá un bloqueo en la licencia con un periodo máximo de 24 horas.

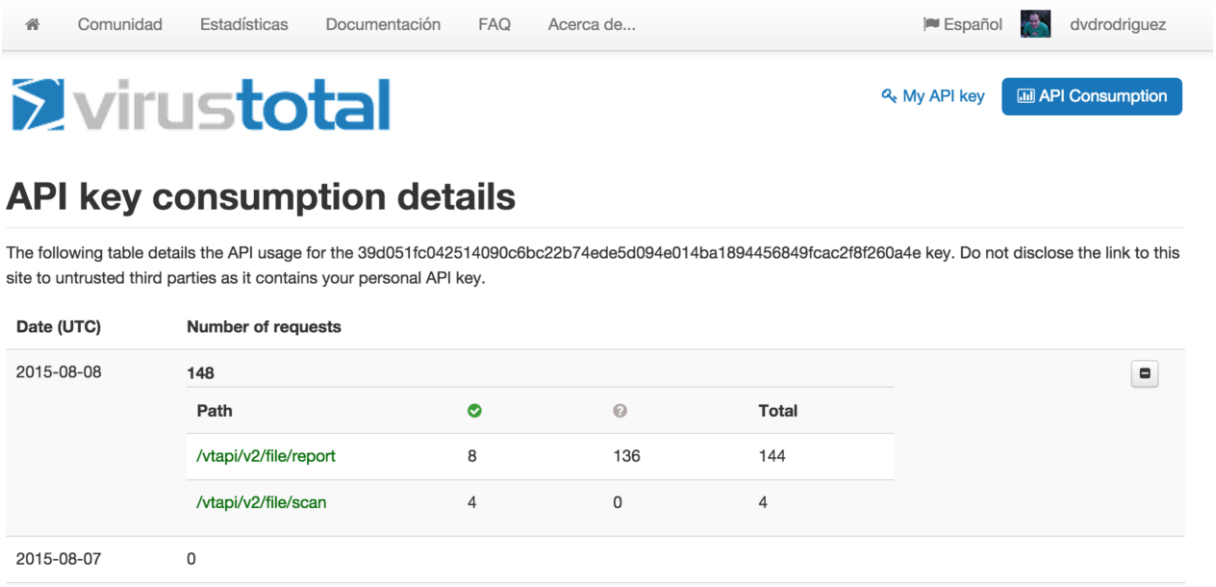


Ilustración 14: Detalles de consumo de la licencia de VirusTotal

5.4 Almacenamiento de los resultados en la base de datos local

En la base de datos local se almacenan una relación de ficheros tratados por máquina, los resultados de los análisis proporcionados por *VirusTotal* y estadísticas sobre la eficacia de los antivirus que han participado en la detección de cada uno de los ficheros sospechosos.

Como se ha comentado en el capítulo 3.3.2 del presente documento, para recopilar toda la información existen tres tablas que se detallan a continuación.

- Tabla "*malware*": En ella se documentan todos los ficheros analizados por *VirusTotal*. En la siguiente imagen se muestra una captura del estado de la tabla después de realizar varios centenares de análisis. Entre los valores almacenados, encontramos:

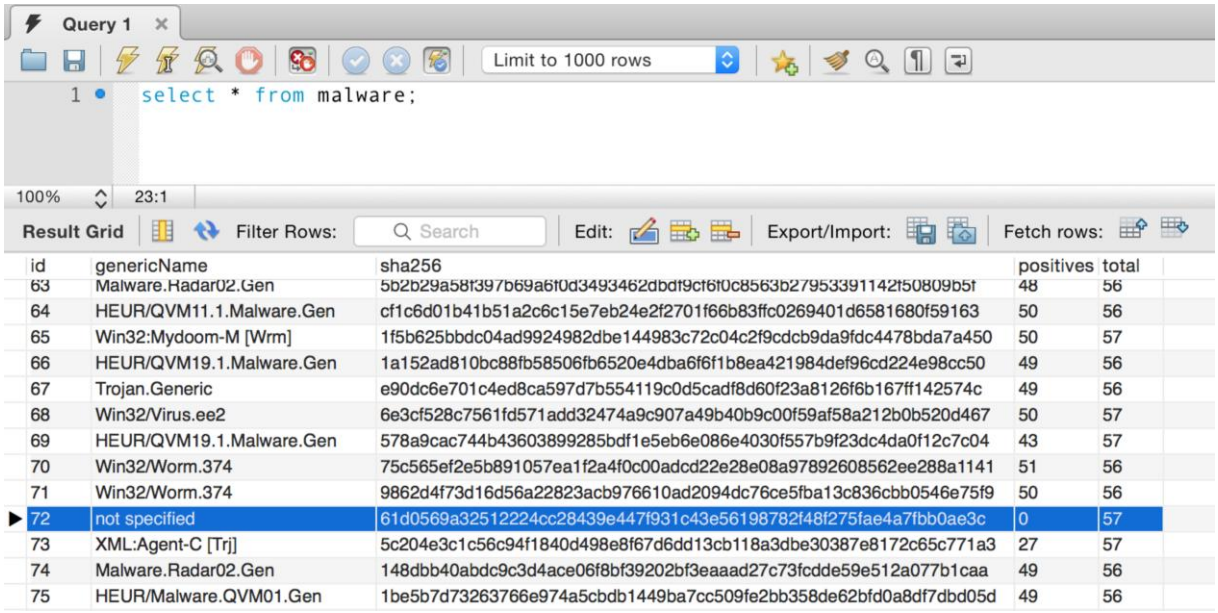
    - Nombre genérico del fichero (campo "*genericName*"): Viene dado por el identificativo que proporcione el resultado del primer antivirus que realice el

análisis. Por tanto, existe la posibilidad que un mismo nombre genérico pueda definir varios ficheros diferentes.

Si no existe malware (es decir, el fichero es considerado como benigno), el nombre aparecerá como "*not specified*".

- SHA-256 (campo "*sha256*"): Se trata del valor obtenido al aplicar la función hash criptográfica SHA-256 sobre el fichero a analizar. Define, a modo interno, la firma del fichero analizado. Es única por cada fichero.
- Número total de antivirus que participan en el análisis (campo "*total*"): Indica cuántos antivirus han participado en el análisis.

  VirusTotal define unos parámetros mínimos para determinar la calidad del servicio. Los análisis deben ser realizados en un tiempo determinado y, en caso de no haber obtenido respuesta dentro de los límites establecidos, se produce un time-out y el antivirus no computa para el análisis. Es por ello por lo que se puede apreciar que la participación de antivirus varía con respecto al fichero analizado.
- Número de positivos (campo "*positives*"): indica el número de antivirus –de entre los totales que han participado en el análisis- que han dado resultado positivo. Esto es, indican que el fichero analizado contiene malware. El valor mínimo varía de entre 0 y el número total de antivirus implicados.

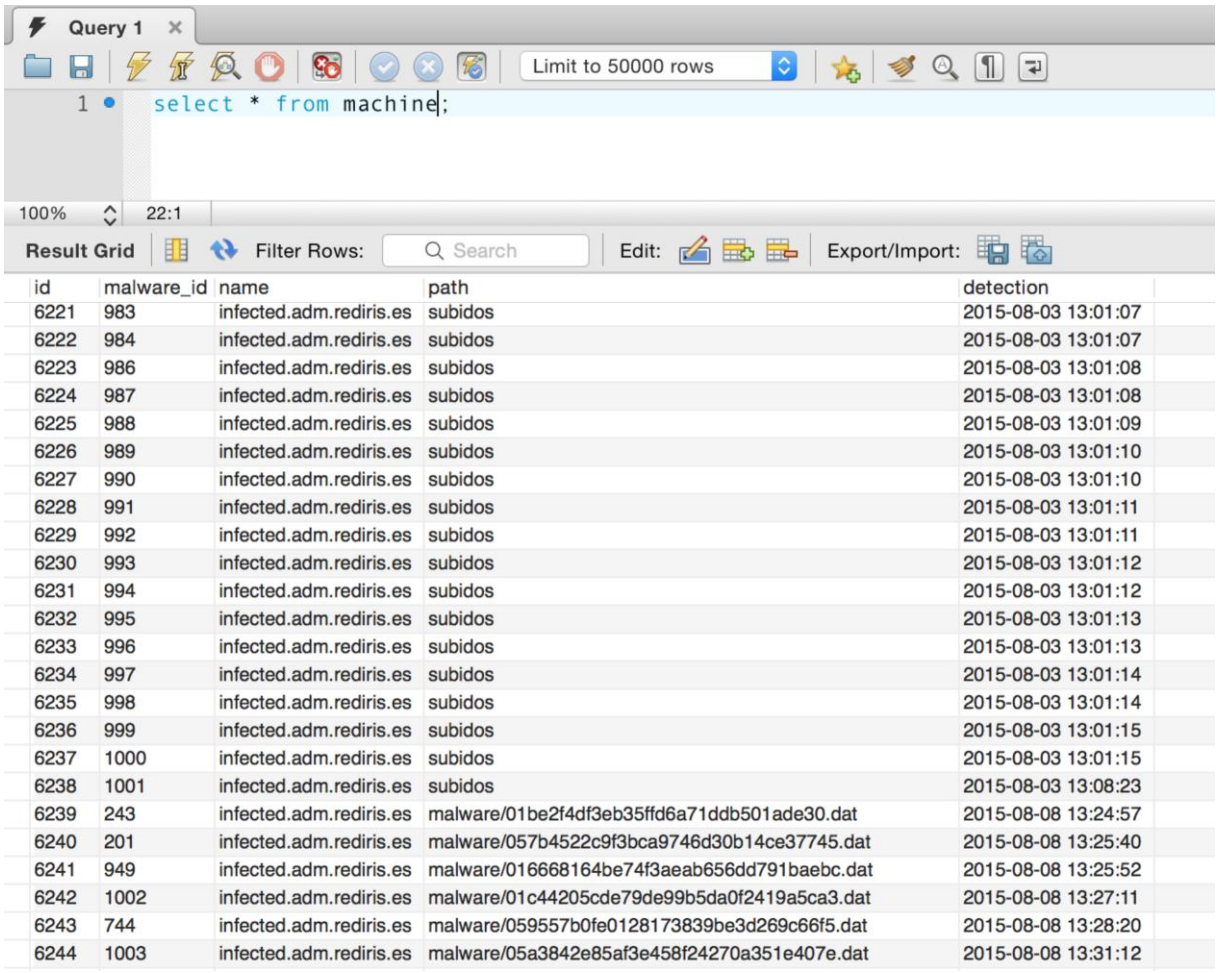


id	genericName	sha256	positives	total
63	Malware.Radar02.Gen	5b2b29a58f397b69a6f0d3493462dbdf9cf6f0c8563b27953391142f50809b5f	48	56
64	HEUR/QVM11.1.Malware.Gen	cf1c6d01b41b51a2c6c15e7eb24e2f2701f66b83ffc0269401d6581680f59163	50	56
65	Win32:Mydoom-M [Wrm]	1f5b625bbdc04ad9924982dbe144983c72c04c2f9cdcb9da9fdc4478bda7a450	50	57
66	HEUR/QVM19.1.Malware.Gen	1a152ad810bc88fb58506fb6520e4dba6f6f1b8ea421984def96cd224e98cc50	49	56
67	Trojan.Generic	e90dc6e701c4ed8ca597d7b554119c0d5cadf8d60f23a8126f6b167ff142574c	49	56
68	Win32/Virus.ee2	6e3cf528c7561fd571add32474a9c907a49b40b9c00f59af58a212b0b520d467	50	57
69	HEUR/QVM19.1.Malware.Gen	578a9cac744b43603899285bdf1e5eb6e086e4030f557b9f23dc4da0f12c7c04	43	57
70	Win32/Worm.374	75c565ef2e5b891057ea1f2a4f0c00adcd22e28e08a97892608562ee288a1141	51	56
71	Win32/Worm.374	9862d4f73d16d56a22823acb976610ad2094dc76ce5fba13c836cbb0546e75f9	50	56
72	not specified	61d0569a32512224cc28439e447f931c43e56198782f48f275fae4a7fbb0ae3c	0	57
73	XML:Agent-C [Trj]	5c204e3c1c56c94f1840d498e8f67d6dd13cb118a3dbe30387e8172c65c771a3	27	57
74	Malware.Radar02.Gen	148dbb40abdc9c3d4ace06f8bf39202bf3eaaad27c73fcdde59e512a077b1caa	49	56
75	HEUR/Malware.QVM01.Gen	1be5b7d73263766e974a5cbdb1449ba7cc509fe2bb358de62bfd0a8df7dbd05d	49	56

Ilustración 15: Almacenamiento de datos en la tabla "*malware*"

- Tabla "*machine*": Mantiene un histórico de todos los ficheros analizados, independientemente que se consideren malware, o no.
    - Columna "*malware_id*": hace referencia a la clave primaria del fichero analizado que se encuentra en la tabla "malware".
    - Columna "*name*": Nombre de la máquina en la que se ha encontrado el fichero. Este atributo se debe encontrar correctamente definido en la librería "*VTMySQL.pm*" del detector de malware.
    - Columna "*path*": Indica la ruta donde se encuentra el fichero analizado en la máquina "*name*". En el caso de análisis de ficheros individuales (no directorios) se indica también el nombre del fichero.

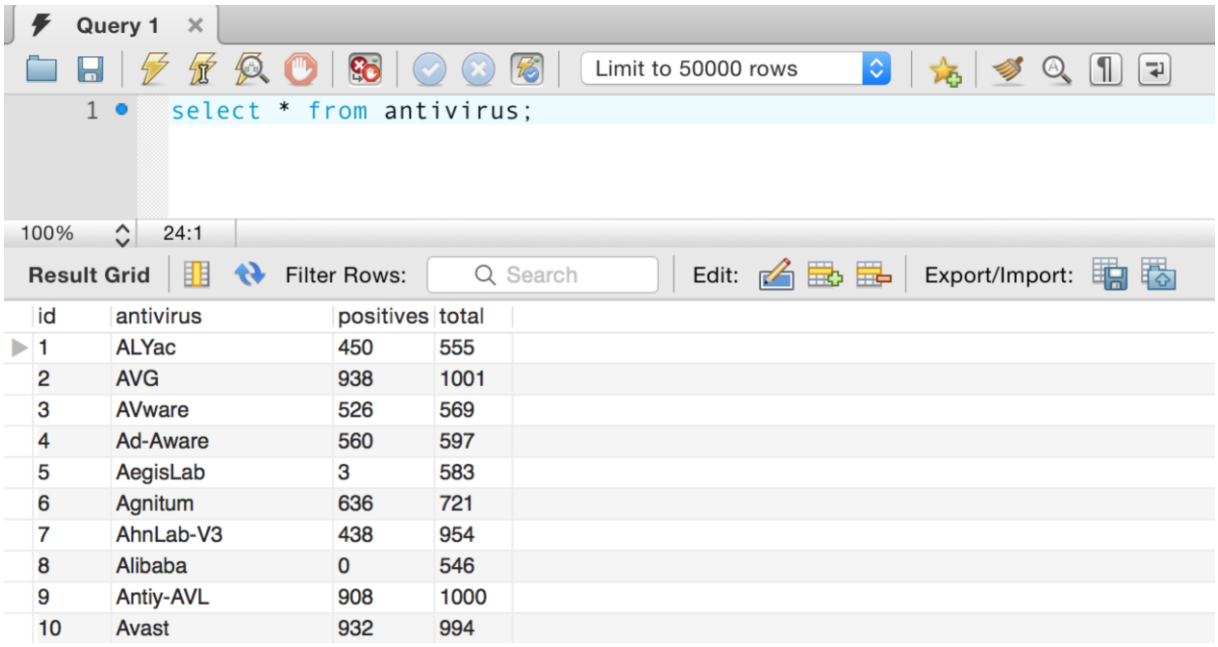
      Se aconseja que en la invocación del programa para análisis de directorios completos se realice con ruta absolutas y no con rutas relativas con respecto a la ubicación del ejecutable del detector de malware. De otro modo, si la ubicación del detector varía, se podría perder la referencia de adónde se encuentran los ficheros maliciosos.
    - Columna "*detection*": Fecha y hora en la que se acometió el análisis del fichero.



id	malware_id	name	path	detection
6221	983	infected.adm.rediris.es	subidos	2015-08-03 13:01:07
6222	984	infected.adm.rediris.es	subidos	2015-08-03 13:01:07
6223	986	infected.adm.rediris.es	subidos	2015-08-03 13:01:08
6224	987	infected.adm.rediris.es	subidos	2015-08-03 13:01:08
6225	988	infected.adm.rediris.es	subidos	2015-08-03 13:01:09
6226	989	infected.adm.rediris.es	subidos	2015-08-03 13:01:10
6227	990	infected.adm.rediris.es	subidos	2015-08-03 13:01:10
6228	991	infected.adm.rediris.es	subidos	2015-08-03 13:01:11
6229	992	infected.adm.rediris.es	subidos	2015-08-03 13:01:11
6230	993	infected.adm.rediris.es	subidos	2015-08-03 13:01:12
6231	994	infected.adm.rediris.es	subidos	2015-08-03 13:01:12
6232	995	infected.adm.rediris.es	subidos	2015-08-03 13:01:13
6233	996	infected.adm.rediris.es	subidos	2015-08-03 13:01:13
6234	997	infected.adm.rediris.es	subidos	2015-08-03 13:01:14
6235	998	infected.adm.rediris.es	subidos	2015-08-03 13:01:14
6236	999	infected.adm.rediris.es	subidos	2015-08-03 13:01:15
6237	1000	infected.adm.rediris.es	subidos	2015-08-03 13:01:15
6238	1001	infected.adm.rediris.es	subidos	2015-08-03 13:08:23
6239	243	infected.adm.rediris.es	malware/01be2f4df3eb35ffd6a71ddb501ade30.dat	2015-08-08 13:24:57
6240	201	infected.adm.rediris.es	malware/057b4522c9f3bca9746d30b14ce37745.dat	2015-08-08 13:25:40
6241	949	infected.adm.rediris.es	malware/016668164be74f3aeab656dd791baebc.dat	2015-08-08 13:25:52
6242	1002	infected.adm.rediris.es	malware/01c44205cde79de99b5da0f2419a5ca3.dat	2015-08-08 13:27:11
6243	744	infected.adm.rediris.es	malware/059557b0fe0128173839be3d269c66f5.dat	2015-08-08 13:28:20
6244	1003	infected.adm.rediris.es	malware/05a3842e85af3e458f24270a351e407e.dat	2015-08-08 13:31:12

Ilustración 16: Almacenamiento de datos en la tabla "*machine*"

- Tabla "*antivirus*": Almacena información con finalidad meramente estadística.
  - Columna "*antivirus*": indica el nombre comercial del fabricante o marca del antivirus.
  - Columna "*total*": número total de ficheros analizados en la organización.
  - Columna "*positives*": refiere a al número de ficheros considerados como malware han sido detectados con respecto al número total de ficheros analizados en la organización. En este caso no se incluyen los ficheros analizados que han sido considerados como benignos.



id	antivirus	positives	total
1	ALYac	450	555
2	AVG	938	1001
3	AVware	526	569
4	Ad-Aware	560	597
5	AegisLab	3	583
6	Agnitum	636	721
7	AhnLab-V3	438	954
8	Alibaba	0	546
9	Antiy-AVL	908	1000
10	Avast	932	994

Ilustración 17: Almacenamiento de datos en la tabla "*antivirus*"

5.5 Visualización vía web

Aunque el servidor web no es requisito *sine qua non* para la ejecución del detector de malware, es altamente recomendable su implantación.

Entre las mejoras, ofrece:

- Visualización de datos sin manipular el SGBD o realizar consultas que puedan comprometer la integridad de los mismos.
- Estadísticas en tiempo real mediante interfaz sencilla, ágil y vistosa.
- Capacidad para filtrar el contenido por usuarios y/o roles.
- Posibilidad de personalización y flexibilidad de desarrollos futuros, así como la integración completa en herramientas corporativas.

Para el desarrollo del Trabajo Fin de Máster, ha sido contratado un hosting con el dominio *www.malwaredetector.net* en un proveedor externo[iv]. De esta forma se puede simular, como si un entorno de producción real se tratase, la funcionalidad completa del detector de malware.

[iv] https://www.vadavo.com/

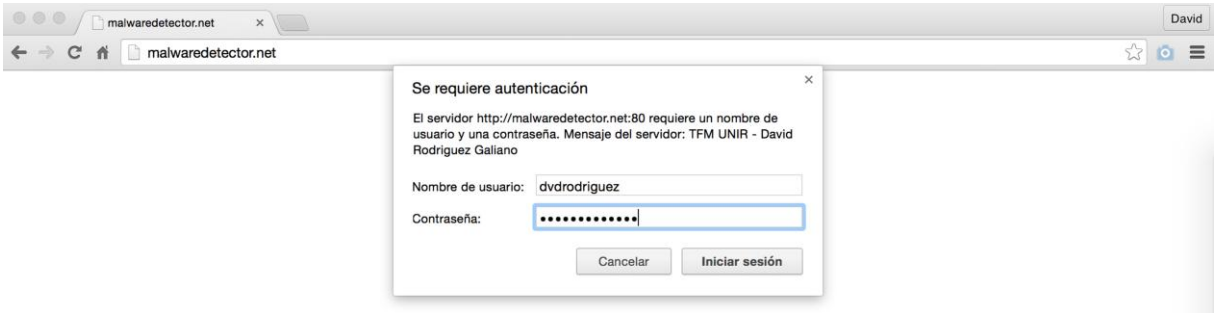


Ilustración 18: Protección de plataforma web para detector de malware

Una vez que se ha accedido a la plataforma web con las debidas credenciales de acceso se visualizan una serie de tablas que cada una de ella ejecuta una consulta sobre las bases de datos.

Para una organización, la funcionalidad más importante es la de poder visualizar qué máquinas están infectadas y dónde se encuentra el malware. Esta funcionalidad queda completamente abarcada en la pestaña "*Detected malware*".

Inicialmente se solicita, mediante un sencillo formulario, sobre qué máquinas se desea realizar la consulta (todas o una específica), con la ventana de tiempo en la que los análisis fueron efectuados (fecha origen y fecha fin). Si no se especifica ningún rango de fechas aparecerán todos los análisis referidos a las máquinas indicadas.

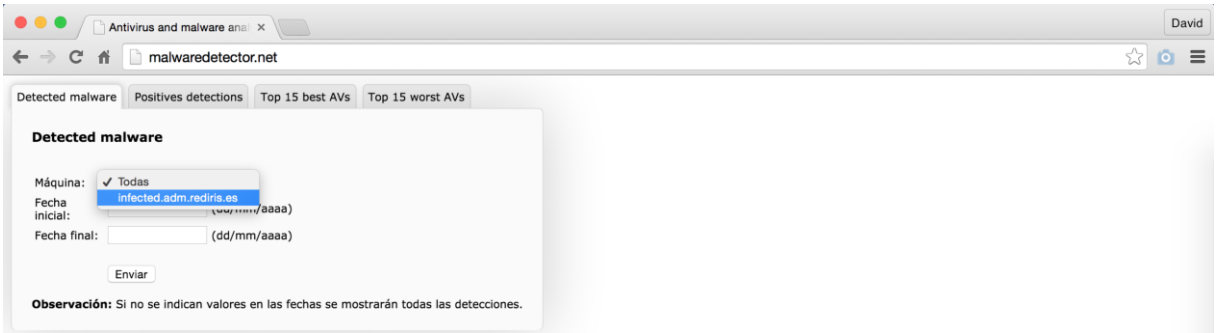
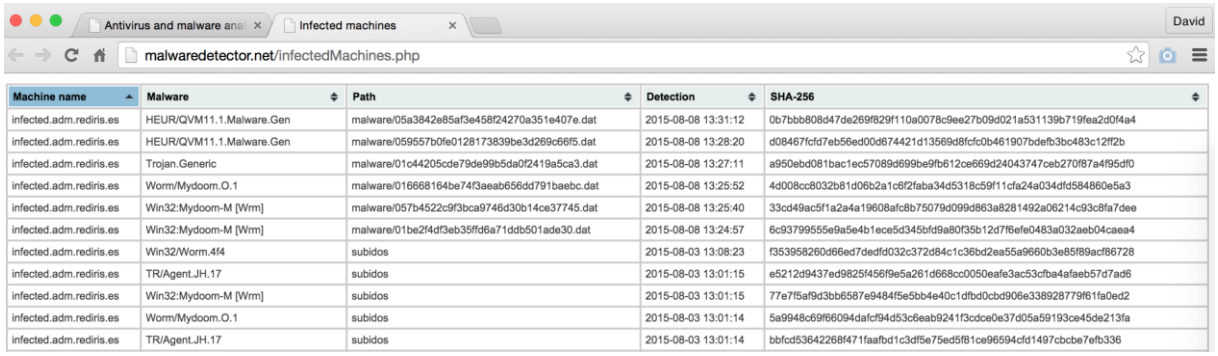


Ilustración 19: Formulario inicial de búsqueda de análisis realizados

Al realizar el *submit* del formulario una nueva pestaña del navegador se abre. En ella aparece la información relevante sobre los ficheros analizados, en qué máquina se encuentran, si contienen o no malware, la ubicación de los mismos, la fecha de detección y, finalmente, la función hash SHA-256 aplicada sobre el fichero.

La tabla permite la ordenación por cada uno de estos atributos en orden ascendente y descendente.



Machine name	Malware	Path	Detection	SHA-256
infected.adm.rediris.es	HEUR/QVM11.1.Malware.Gen	malware/05a3842e85af3e458f24270a351e407e.dat	2015-08-08 13:31:12	0b7bbb808d47de269f829f110a0078c9ee27b09d021a531139b719fea2d0f4a4
infected.adm.rediris.es	HEUR/QVM11.1.Malware.Gen	malware/059557b0fe0128173839be3d269c66f5.dat	2015-08-08 13:28:20	d08467fcfd7eb56ed00d674421d13569d8fcfc0b461907bdefb3bc483c12ff2b
infected.adm.rediris.es	Trojan.Generic	malware/01c44205cde79de99b5da0f2419a5ca3.dat	2015-08-08 13:27:11	a950ebd081bac1ec57089d699be9fb612ce669d24043747ceb270f87a4f95df0
infected.adm.rediris.es	Worm/Mydoom.O.1	malware/016668164be74f3aeab656dd791baebc.dat	2015-08-08 13:25:52	4d008cc8032b81d06b2a1c6f2faba34d5318c59f11cfa24a034dfd584860e5a3
infected.adm.rediris.es	Win32:Mydoom-M [Wrm]	malware/057b4522c9f3bca9746d30b14ce37745.dat	2015-08-08 13:25:40	33cd49ac5f1a2a4a19608afc8b75079d099d863a8281492a06214c93c8fa7dee
infected.adm.rediris.es	Win32:Mydoom-M [Wrm]	malware/01be2f4df3eb35ffd6a71ddb501ade30.dat	2015-08-08 13:24:57	6c93799555e9a5e4b1ece5d345bfd9a80f35b12d7f6efe0483a032aeb04caea4
infected.adm.rediris.es	Win32/Worm.4f4	subidos	2015-08-03 13:08:23	f353958260d66ed7dedfd032c372d84c1c36bd2ea55a9660b3e85f89acf86728
infected.adm.rediris.es	TR/Agent.JH.17	subidos	2015-08-03 13:01:15	e5212d9437ed9825f456f9e5a261d668cc0050eafe3ac53cfba4afaeb57d7ad6
infected.adm.rediris.es	Win32:Mydoom-M [Wrm]	subidos	2015-08-03 13:01:15	77e7f5af9d3bb6587e9484f5e5bb4e40c1dfbd0cbd906e338928779f61fa0ed2
infected.adm.rediris.es	Worm/Mydoom.O.1	subidos	2015-08-03 13:01:14	5a9948c69f66094dafcf94d53c6eab9241f3cdce0e37d05a59193ce45de213fa
infected.adm.rediris.es	TR/Agent.JH.17	subidos	2015-08-03 13:01:14	bbfcd53642268f471faafbd1c3df5e75ed5f81ce96594cfd1497cbcbe7efb336

Ilustración 20: Relación de análisis realizados

Adicionalmente, para verificar la flexibilidad de la aplicación en su personalización y capacidad de crecimiento futuro, se muestra un ejemplo de análisis estadístico incluido en la plataforma web. Como se puede apreciar, permite la incorporación de software libre para completar la funcionalidad deseada, como la herramienta de gráficas mediante JavaScript *amCharts*[v].

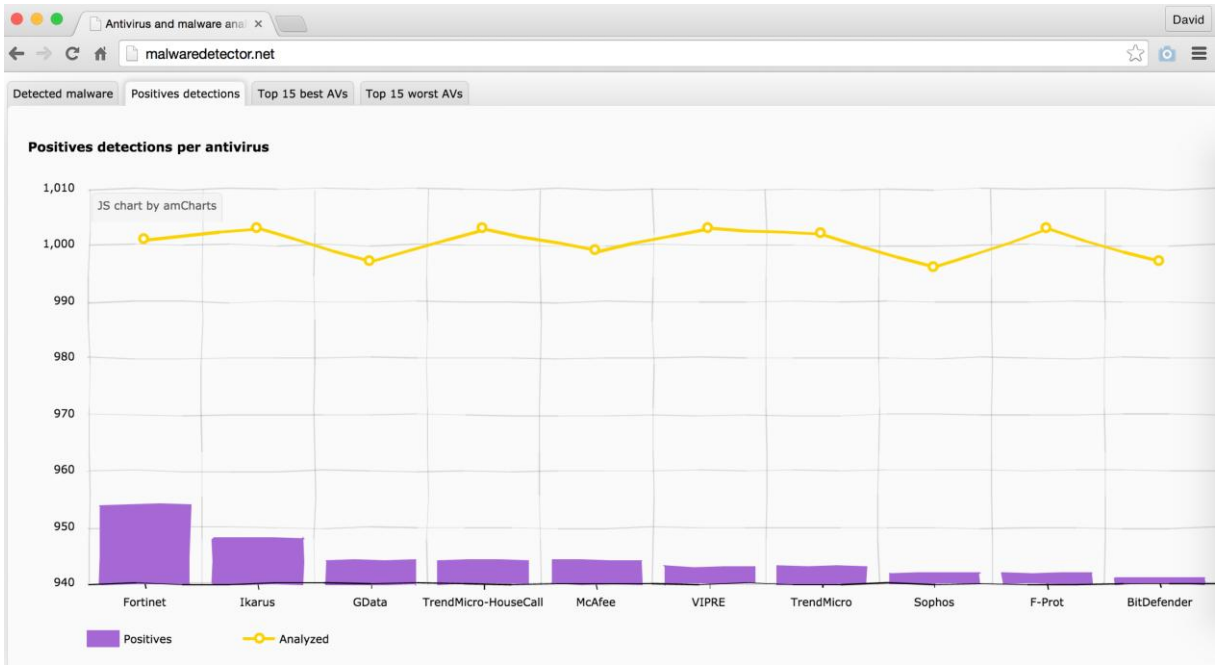


Ilustración 21: Ejemplo de análisis estadístico en plataforma web del detector de malware

[v] http://www.amcharts.com/

Capítulo 6 Modelados de datos y análisis de antivirus

Fruto del desarrollo del detector de malware multiplataforma en infraestructuras descentralizadas, se ha podido realizar un estudio sobre la eficacia de los antivirus que participan en el proyecto *VirusTotal*, evaluando única y exclusivamente la capacidad de éstos para encontrar malware. Es decir, no se contempla (pues no es la finalidad de este Trabajo Fin de Máster) la evaluación referida a penalización en el rendimiento o ralentización de las máquinas en las que se han instalado los antivirus a analizar.

Se pretende conocer, dentro de la amplia gama de antivirus gratuitos y comerciales cuál se puede considerar más confiable en el desempeño de su labor.

Para ello, la Red Académica y de Investigación Española (RedIRIS)[vi] facilitó ficheros maliciosos que han sido detectados y analizados en universidades y centros públicos de investigación.

La prueba ha consistido en analizar 1000 ficheros que, a ciencia cierta, se consideran maliciosos; bien porque son conocidos y su clasificación se ha hecho pública o bien porque en el análisis forense de máquinas infectadas se ha detectado que generan un comportamiento anómalo y/o no deseado.

[10] Cuando hablo de malware, me refiero a virus clásicos, gusanos de red, caballos de troya, spyware, phishing, adware, riskware, bromas, rootkits, spammers o cualquier otro programa que no afecta directamente a los ordenadores, pero que se usan para crear virus, troyanos o para realizar actividades ilegales como ataques DoS y penetrar en otros equipos.

En los siguientes apartados se muestran los resultados obtenidos.

6.1 Número total de ficheros analizados

De los 67 antivirus que han participado en los análisis, sólo 7 han analizado los 1000 ficheros maliciosos subidos al proyecto *VirusTotal*.

[vi] http://www.rediris.es/

El motivo por el que no todos los antivirus han participado en el análisis del 100% de los ficheros subidos se debe, según los ingenieros de *VirusTotal*, a que han proporcionado un time-out por haber tardado demasiado en escanear y no fueron capaces de producir un resultado en el umbral de tiempo que tienen estipulado.

Aunque, como se ha comentado, no se tiene en cuenta la eficiencia del antivirus en el proceso de análisis, sí que se puede percibir cuáles tienen más dificultad para realizar un análisis completo en tiempos medianamente aceptables.

Bien es cierto que las versiones de los antivirus que participan[11] en el proyecto *VirusTotal* no tienen porqué ser iguales que las propias versiones comerciales. Los fabricantes pueden parametrizar los motores del antivirus personalizados para *VirusTotal*; por ejemplo, aplicando una heurística más fuerte o inclusión de firmas beta.

A pesar de ello, por temas de competencia e imagen, los fabricantes de antivirus no expondrán sus peores galas en *VirusTotal*, por lo que los resultados obtenidos permiten tener una idea de cuán buena es su solución.

Tabla 4: Datos de ficheros analizados por antivirus con respecto al total (1000)

Antivirus	Número total de ficheros analizados con respecto a 1000	% Ficheros analizados sobre 1000
CAT-QuickHeal	1000	100,00%
F-Prot	1000	100,00%
Ikarus	1000	100,00%
SUPERAntiSpyware	1000	100,00%
TheHacker	1000	100,00%
TrendMicro-HouseCall	1000	100,00%
VIPRE	1000	100,00%
ClamAV	999	99,90%
TrendMicro	999	99,90%
AVG	998	99,80%
Fortinet	998	99,80%
Antiy-AVL	997	99,70%
K7AntiVirus	996	99,60%
McAfee	996	99,60%
McAfee-GW-Edition	996	99,60%
Panda	996	99,60%

Kaspersky	995	99,50%
BitDefender	994	99,40%
GData	994	99,40%
Microsoft	994	99,40%
Jiangmin	993	99,30%
Sophos	993	99,30%
ViRobot	992	99,20%
Avast	991	99,10%
nProtect	991	99,10%
VBA32	990	99,00%
Comodo	989	98,90%
DrWeb	989	98,90%
Rising	989	98,90%
TotalDefense	985	98,50%
ByteHero	983	98,30%
Symantec	983	98,30%
Emsisoft	981	98,10%
F-Secure	980	98,00%
AhnLab-V3	951	95,10%
ESET-NOD32	877	87,70%
Agnitum	718	71,80%
Kingsoft	707	70,70%
MicroWorld-eScan	702	70,20%
NANO-Antivirus	617	61,70%
K7GW	613	61,30%
Malwarebytes	609	60,90%
Baidu-International	604	60,40%
Norman	595	59,50%
Ad-Aware	594	59,40%
Qihoo-360	586	58,60%
Bkav	581	58,10%
AegisLab	580	58,00%
Zoner	576	57,60%
Zillya	573	57,30%
Cyren	569	56,90%
Tencent	569	56,90%

AVware	566	56,60%
ALYac	552	55,20%
Alibaba	543	54,30%
Avira	506	50,60%
AntiVir	432	43,20%
Commtouch	431	43,10%
Arcabit	401	40,10%
eSafe	384	38,40%
PCTools	379	37,90%
VirusBuster	276	27,60%
CMC	188	18,80%
NOD32	117	11,70%
eTrust-Vet	8	0,80%
Prevx	3	0,30%
eScan	3	0,30%

En la siguiente gráfica se muestra mediante un diagrama de barras los 20 mejores antivirus si nos referimos en exclusividad al número de ficheros analizados.

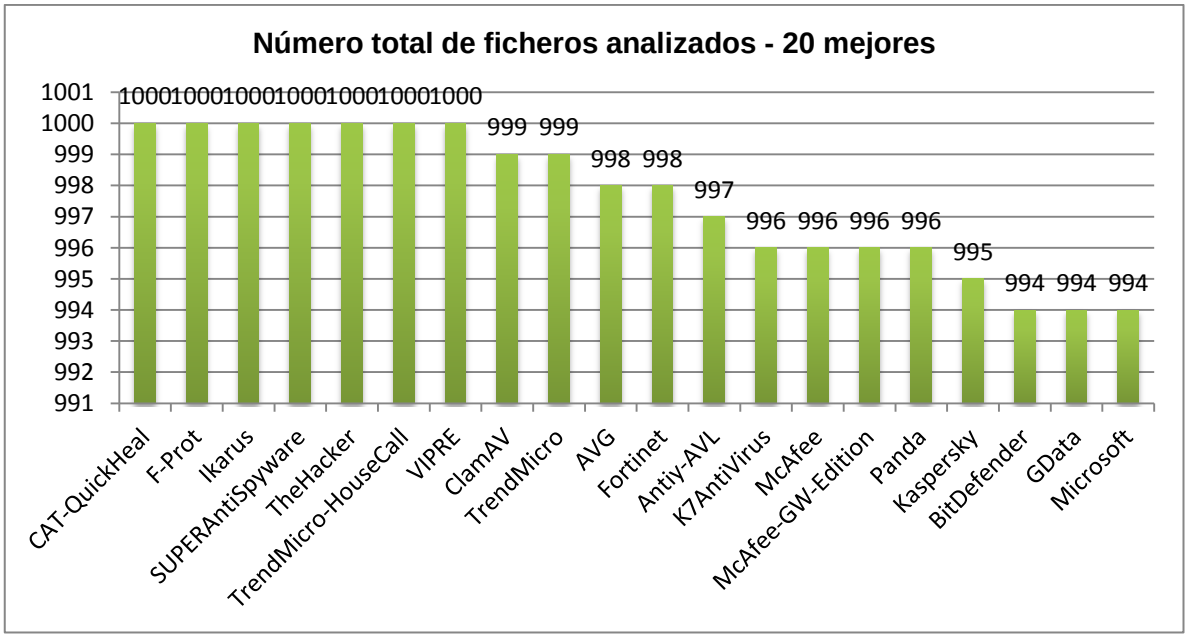


Ilustración 22: 20 mejores antivirus por número total de ficheros analizados

De igual forma, los 20 peores antivirus son los que se muestran a continuación. Fijémonos que 11 antivirus analizan menos de la mitad de los archivos subidos. Especialmente es

curioso el dato referido a NOD32 (considerado popularmente como uno de los mejores antivirus) y que vagamente analiza poco más de un 11% de los ficheros subidos.

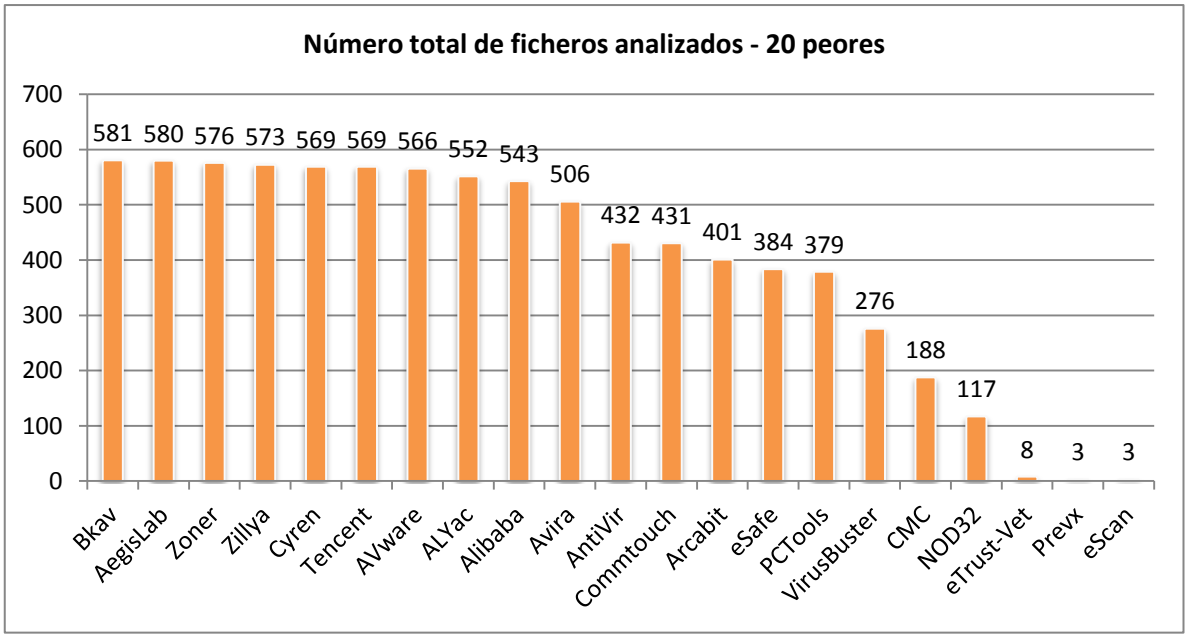


Ilustración 23: 20 peores antivirus por número total de ficheros analizados

6.2 Número de positivos detectados por antivirus

En la siguiente tabla (ordenados de forma descendente por la columna "*Ficheros analizados considerados positivos*") se muestran los resultados fruto de la evaluación de cada uno de los antivirus referido al número de ficheros considerados realmente como malware. Recordemos que los 1000 ficheros subidos al proyecto *VirusTotal* son maliciosos, sin embargo no todos los antivirus los consideran como tal.

Como se puede observar, solo "*eScan*" ha considerado el 100% de los ficheros analizados como maliciosos. Ahora bien, sólo escaneó 3; es decir el 0,3% del total.

A mayor sorpresa, nos encontramos antivirus que no han considerado ninguno de los archivos analizados como malware. Es el caso de "*Alibaba*" y "*Prevx*".

Tabla 5: Datos de ficheros considerados como positivos por antivirus

Antivirus	Ficheros analizados considerados positivos	Número total de ficheros analizados con respecto a 1000	% Ficheros considerados positivos con respecto a los ficheros analizados	% Ficheros considerados positivos con respecto al total de ficheros subidos (1000)
Fortinet	952	998	95,39%	95,20%
Ikarus	946	1000	94,60%	94,60%
GData	942	994	94,77%	94,20%
McAfee	942	996	94,58%	94,20%
TrendMicro-HouseCall	942	1000	94,20%	94,20%
TrendMicro	941	999	94,19%	94,10%
VIPRE	941	1000	94,10%	94,10%
F-Prot	940	1000	94,00%	94,00%
Sophos	940	993	94,66%	94,00%
BitDefender	939	994	94,47%	93,90%
Kaspersky	937	995	94,17%	93,70%
AVG	936	998	93,79%	93,60%
McAfee-GW-Edition	931	996	93,47%	93,10%
Avast	930	991	93,84%	93,00%
Microsoft	928	994	93,36%	92,80%
CAT-QuickHeal	925	1000	92,50%	92,50%
Comodo	925	989	93,53%	92,50%
F-Secure	924	980	94,29%	92,40%
DrWeb	917	989	92,72%	91,70%
Panda	917	996	92,07%	91,70%
Symantec	917	983	93,29%	91,70%
Emsisoft	914	981	93,17%	91,40%
VBA32	910	990	91,92%	91,00%
nProtect	907	991	91,52%	90,70%
Antiy-AVL	906	997	90,87%	90,60%
K7AntiVirus	905	996	90,86%	90,50%
TheHacker	895	1000	89,50%	89,50%
ClamAV	892	999	89,29%	89,20%

Rising	892	989	90,19%	89,20%
Jiangmin	890	993	89,63%	89,00%
TotalDefense	885	985	89,85%	88,50%
ViRobot	864	992	87,10%	86,40%
ESET-NOD32	823	877	93,84%	82,30%
Agnitum	634	718	88,30%	63,40%
MicroWorld-eScan	633	702	90,17%	63,30%
Kingsoft	604	707	85,43%	60,40%
NANO-Antivirus	562	617	91,09%	56,20%
Ad-Aware	558	594	93,94%	55,80%
Norman	555	595	93,28%	55,50%
Malwarebytes	553	609	90,80%	55,30%
Cyren	531	569	93,32%	53,10%
AVware	524	566	92,58%	52,40%
Zillya	488	573	85,17%	48,80%
K7GW	484	613	78,96%	48,40%
Bkav	477	581	82,10%	47,70%
Avira	461	506	91,11%	46,10%
Zoner	449	576	77,95%	44,90%
ALYac	448	552	81,16%	44,80%
AhnLab-V3	438	951	46,06%	43,80%
AntiVir	415	432	96,06%	41,50%
Commtouch	409	431	94,90%	40,90%
Arcabit	372	401	92,77%	37,20%
eSafe	351	384	91,41%	35,10%
PCTools	341	379	89,97%	34,10%
Tencent	316	569	55,54%	31,60%
Qihoo-360	299	586	51,02%	29,90%
VirusBuster	267	276	96,74%	26,70%
Baidu-International	231	604	38,25%	23,10%
SUPERAntiSpyware	156	1000	15,60%	15,60%
CMC	131	188	69,68%	13,10%

NOD32	115	117	98,29%	11,50%
ByteHero	21	983	2,14%	2,10%
eTrust-Vet	5	8	62,50%	0,50%
AegisLab	3	580	0,52%	0,30%
eScan	3	3	100,00%	0,30%
Alibaba	0	543	0,00%	0,00%
Prevx	0	3	0,00%	0,00%

Con respecto a los 1000 ficheros malware subidos, "*Fortinet*" es el antivirus que más ficheros detecta como maliciosos.

De entre los 20 mejores antivirus existe un equilibrio entre el número de ficheros analizados y la eficacia de ellos.

A continuación se muestran dos gráficas que relacionan el número de ficheros analizados por cada antivirus con el número de detecciones de malware efectuadas. En ellas se indican, en primera instancia, los 20 mejores antivirus y los 20 peores, respectivamente.

Leyenda:

- Color naranja: Número de ficheros analizados por el antivirus
- Color rojo: Número de positivos detectados

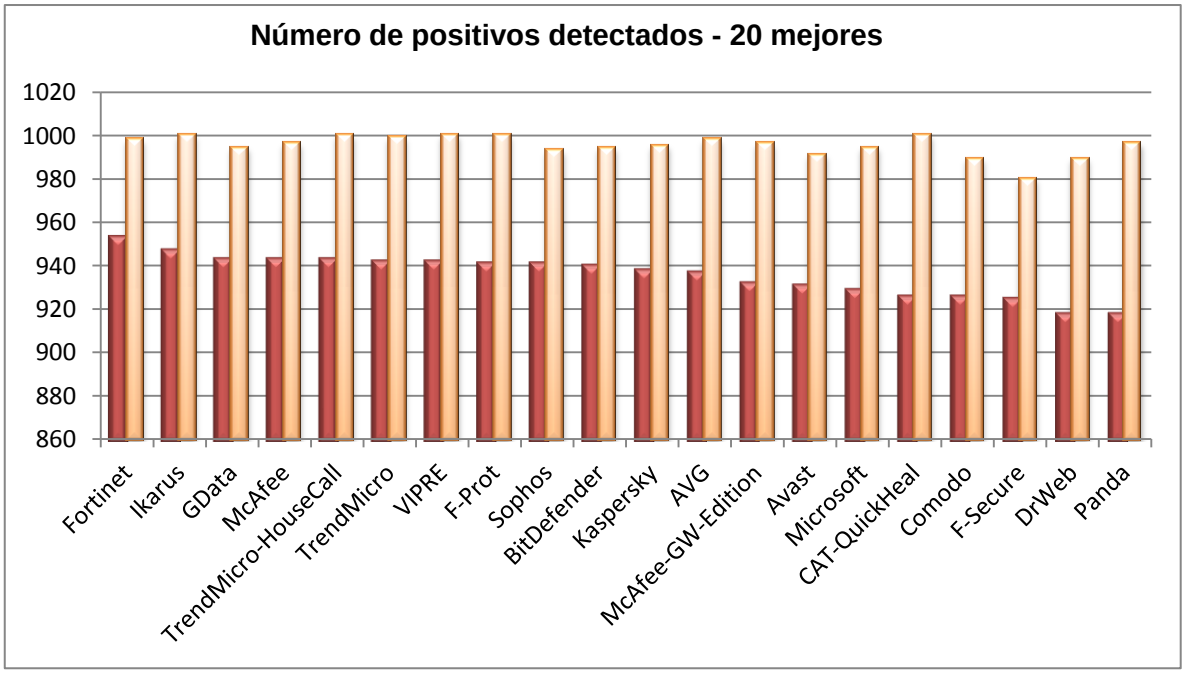


Ilustración 24: 20 mejores antivirus por número total de malware detectado

Cabe destacar que existen antivirus que no han detectado ningún fichero malicioso. En la gráfica siguientes se hace referencia a los 20 peores antivirus por número de detecciones realizadas.

Leyenda:

- Color verde: Número de ficheros analizados por el antivirus
- Color violeta: Número de positivos detectados

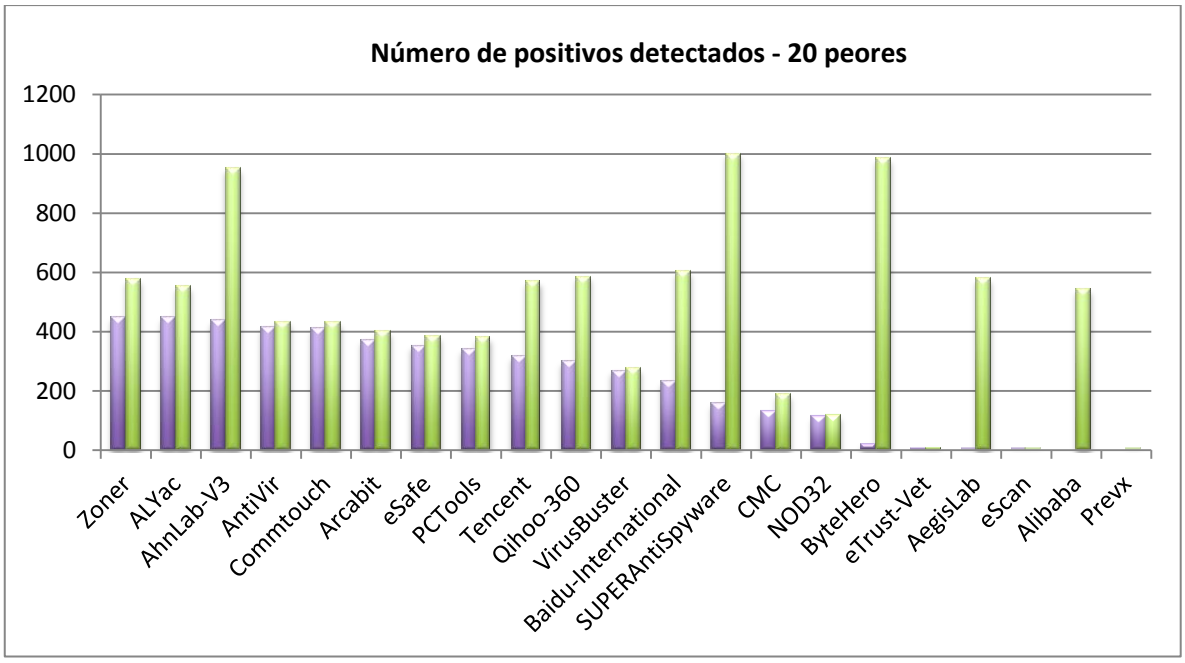


Ilustración 25: 20 peores antivirus por número total de malware detectado

6.3 Relación de falsos negativos

Un falso negativo es un fichero que, a sabiendas de conocer que es malware, es considerado como legítimo e inofensivo. Los falsos negativos son comunes en antivirus basados únicamente en firmas[12], ya que al aparecer virus nuevos o modificados son desconocidos y es necesario una protección adicional.

En caso contrario, encontramos los falsos positivos, que son ficheros benignos que han sido detectados como malware. Los falsos positivos suelen producirse con programas que tienen características de comportamiento similares a los virus, aunque en realidad no lo son.

	Antivirus detected it	Antivirus didn't detect it
It's a virus!	TRUE POSITIVE (caught the virus)	*FALSE NEGATIVE (virus slipped past)*
It's not a virus	*FALSE POSITIVE (quarantined a valid file)*	TRUE NEGATIVE (left well enough alone)

Ilustración 26: Clasificación de ficheros analizados por tipo de resultado (http://www.pcmag.com/article2/0,2817,2481367,00.asp)

En las pruebas que se han realizado sobre el detector de malware se han evaluado en exclusividad los falsos negativos al considerarse los más peligrosos para un sistema informático.

El motivo de ello, es que un falso positivo puede ser detectado como tal mediante un análisis con otro escáner o la propia investigación mediante un análisis forense por parte de ingenieros o expertos en seguridad.

Sin embargo, un falso negativo son ficheros completamente confiables para los antivirus que realmente están causando un mal a la organización.

Tal como se muestra en la siguiente gráfica si el número de falsos positivos aumenta, el número de falsos negativos disminuye. De hecho, ésta es una técnica que comúnmente utilizan frecuentes fabricantes de antivirus con el fin de disminuir el número de falsos negativos.

Lo ideal es alcanzar un equilibrio entre el número de falsos positivos y el de falsos negativos. Este punto se encuentra en la intersección de ambas tasas.

Adicionalmente, dos conclusiones que se alcanzan del estudio de la gráfica, son[13]:

- A mayor sensibilidad del antivirus, mayor posibilidad de detección de falsos positivos y menos aparición de falsos negativos.
- A menor sensibilidad del antivirus, menor detección de falsos positivos y mayor aparición de falsos negativos.

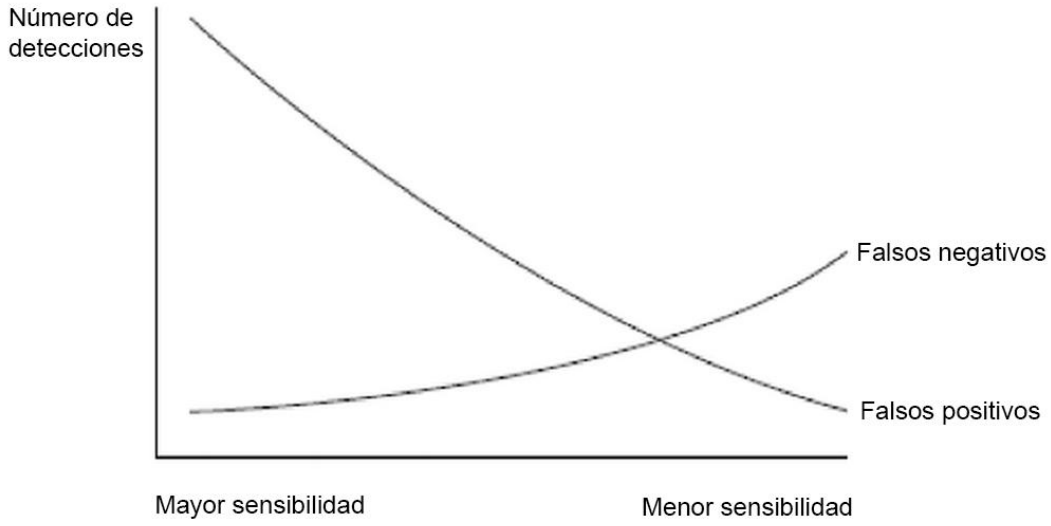


Ilustración 27: Relación entre falsos positivos y falsos negativos.
(Gestión de incidentes en seguridad informática. IFCT0109. Chicano Tejada, Esther. 2014)

Si nos centramos en las pruebas realizadas con el detector de malware y el uso de la fuente de conocimiento de *VirusTotal*, tomaremos el número de falsos positivos como la diferencia entre los ficheros analizados por un antivirus y aquellos considerados como positivos. Es decir, es necesario conocer de entre todos los ficheros subidos (a sabiendas que son maliciosos) cuántos de ellos no han sido detectados por cada antivirus.

Es muy importante tener en cuenta que no se consideran falsos positivos aquellos ficheros que no han sido analizados, puesto que no ha habido un análisis previo y, por tanto, no hay involucración de la toma de decisión del antivirus.

En la gráfica siguiente se puede comprobar, a modo de ejemplo, que el antivirus denominado "*eScan*" sólo analiza 3 ficheros de los 1000 subidos a la plataforma. A pesar de ello, los tres ficheros analizados son detectados como maliciosos. Por tanto, "*eScan*" proporciona un 0% de falsos negativos.

Tabla 6: Porcentaje de falsos negativos por antivirus

Antivirus	Ficheros analizados considerados positivos	Número total de ficheros analizados con respecto a 1000	% Ficheros analizados sobre 1000	% Falsos negativos
Ikarus	946	1000	100,00%	5,40%
TrendMicro-HouseCall	942	1000	100,00%	5,80%
VIPRE	941	1000	100,00%	5,90%
F-Prot	940	1000	100,00%	6,00%
CAT-QuickHeal	925	1000	100,00%	7,50%
TheHacker	895	1000	100,00%	10,50%
SUPERAntiSpyware	156	1000	100,00%	84,40%
TrendMicro	941	999	99,90%	5,80%
ClamAV	892	999	99,90%	10,70%
Fortinet	952	998	99,80%	4,60%
AVG	936	998	99,80%	6,20%
Antiy-AVL	906	997	99,70%	9,10%
McAfee	942	996	99,60%	5,40%
McAfee-GW-Edition	931	996	99,60%	6,50%
Panda	917	996	99,60%	7,90%
K7AntiVirus	905	996	99,60%	9,10%
Kaspersky	937	995	99,50%	5,80%
GData	942	994	99,40%	5,20%
BitDefender	939	994	99,40%	5,50%
Microsoft	928	994	99,40%	6,60%
Sophos	940	993	99,30%	5,30%
Jiangmin	890	993	99,30%	10,30%
ViRobot	864	992	99,20%	12,80%
Avast	930	991	99,10%	6,10%

nProtect	907	991	99,10%	8,40%
VBA32	910	990	99,00%	8,00%
Comodo	925	989	98,90%	6,40%
DrWeb	917	989	98,90%	7,20%
Rising	892	989	98,90%	9,70%
TotalDefense	885	985	98,50%	10,00%
Symantec	917	983	98,30%	6,60%
ByteHero	21	983	98,30%	96,20%
Emsisoft	914	981	98,10%	6,70%
F-Secure	924	980	98,00%	5,60%
AhnLab-V3	438	951	95,10%	51,30%
ESET-NOD32	823	877	87,70%	5,40%
Agnitum	634	718	71,80%	8,40%
Kingsoft	604	707	70,70%	10,30%
MicroWorld-eScan	633	702	70,20%	6,90%
NANO-Antivirus	562	617	61,70%	5,50%
K7GW	484	613	61,30%	12,90%
Malwarebytes	553	609	60,90%	5,60%
Baidu-International	231	604	60,40%	37,30%
Norman	555	595	59,50%	4,00%
Ad-Aware	558	594	59,40%	3,60%
Qihoo-360	299	586	58,60%	28,70%
Bkav	477	581	58,10%	10,40%
AegisLab	3	580	58,00%	57,70%
Zoner	449	576	57,60%	12,70%
Zillya	488	573	57,30%	8,50%
Cyren	531	569	56,90%	3,80%
Tencent	316	569	56,90%	25,30%
AVware	524	566	56,60%	4,20%
ALYac	448	552	55,20%	10,40%

Alibaba	0	543	54,30%	54,30%
Avira	461	506	50,60%	4,50%
AntiVir	415	432	43,20%	1,70%
Commtouch	409	431	43,10%	2,20%
Arcabit	372	401	40,10%	2,90%
eSafe	351	384	38,40%	3,30%
PCTools	341	379	37,90%	3,80%
VirusBuster	267	276	27,60%	0,90%
CMC	131	188	18,80%	5,70%
NOD32	115	117	11,70%	0,20%
eTrust-Vet	5	8	0,80%	0,30%
eScan	3	3	0,30%	0,00%
Prevx	0	3	0,30%	0,30%

Leyenda:

- Color rojo: Porcentaje de ficheros analizados con respecto al total (1000)
- Color azul: Porcentaje de falsos positivos con respecto al número de ficheros analizados por cada antivirus

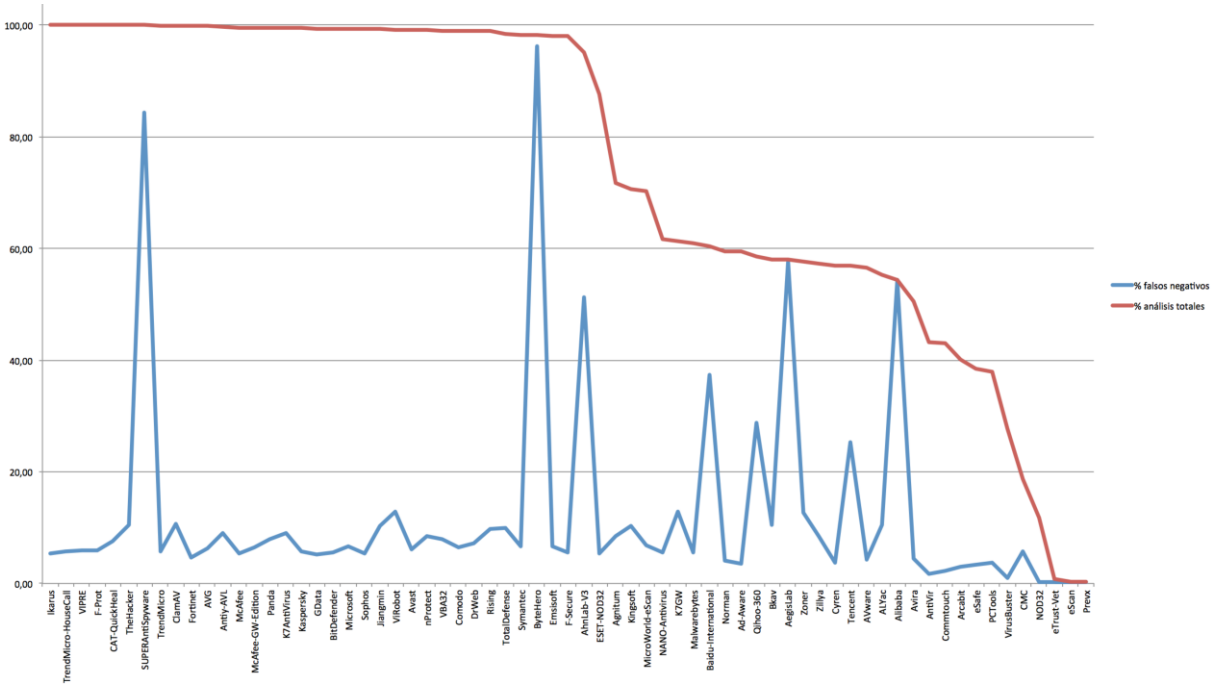


Ilustración 28: Relación de falsos positivos por antivirus

Capítulo 7 Incidencias en el uso del servicio VirusTotal

En el desarrollo del detector de malware multiplataforma en infraestructuras descentralizadas se han localizado una serie de inconvenientes que, a pesar de no estar documentadas por los ingenieros del proyecto *VirusTotal*, sí se ha percibido que la funcionalidad varía dependiendo de ciertos elementos y, por tanto, afecta a la eficiencia del software implementado. No así al resultado del proceso de análisis de cada antivirus, pues éstos permanecen invariables independientemente de las condiciones de ejecución.

La ejecución del detector de malware se ha realizado en múltiples máquinas de diferente naturaleza: máquinas físicas y máquinas virtuales con sistema operativo Linux y distribuciones Debian (Ubuntu), Red Hat Enterprise Linux y CentOS.

Entre los elementos que afectan a la funcionalidad (tiempos de respuesta, códigos de error y time-out), son:

- En redes ultrarrápidas (como fibra oscura) los tiempos de respuesta disminuyen considerablemente con respecto a consultas realizadas sobre líneas de ADSL ordinarias.
- Es realmente complicado establecer el tiempo de espera de entre la consulta y tiempo de respuesta por parte de *VirusTotal*, pues varía en función de si el fichero a analizar ya se encuentra en el repositorio del proyecto, si existe y debe ser reanalizado o, si por el contrario, debe ser subido por primera vez y esperar los resultados de todos los análisis (teniendo en cuenta que los análisis se realizan de forma asíncrona entre los distintos fabricantes de antivirus).

  Adicionalmente a la explicación anterior, también se encuentran las políticas de ejecución que afectan directamente al tiempo de respuesta (como número de peticiones encoladas, disponibilidad y saturación del sistema de *VirusTotal*, etc.). Por tanto, un alto porcentaje de respuestas se pierden por desconocer cuáles deben ser exactamente los tiempos de espera.

  Como medida preventiva, el detector de malware desarrollado espera entre la petición y la respuesta un máximo de 15 segundos. De esta forma se evita una demora no asumible para análisis realizados en masa.
- A pesar de disponer de licencia académica, en el que el número de peticiones es de 20000 por día, éstas se consumen relativamente rápido, pues el número de errores

de cómputo y/o time-out es bastante elevado (siendo considerablemente altos en líneas ADSL).

Bien es cierto, que el detector de malware actúa de manera agresiva; es decir, ante la obtención de un time-out (no error del sistema), se realiza de nuevo la petición (pasado un tiempo prudencial establecido en 180 segundos), consumiendo una solicitud más de entre las proporcionadas en la licencia.

La base de datos local, a medida que se introducen los datos proporcionados por *VirusTotal*, minimiza el número de peticiones realizadas que se pierden. Por tanto, fruto de acumular experiencia, la eficiencia de las consultas cada vez son mejores.

Capítulo 8 Objetivos logrados y trabajo futuro

El desarrollo del detector de malware permite la identificación de ficheros maliciosos en infraestructuras multiplataforma y descentralizadas. Por tanto, facilita la asistencia sobre posibles máquinas comprometidas de los ingenieros de seguridad de cualquier organización.

Cada fabricante de antivirus dispone de sus firmas para la detección, análisis y –en algunos casos- desinfección, pero al existir múltiples plataformas que unifican la funcionalidad (como *VirusTotal*, *Metascan*, etc.) es recomendable unificar esfuerzos para disponer de la mayor fuente de conocimiento posible y hacer de la lucha contra el malware una actividad coordinada, eficaz y eficiente.

Recordemos que un único antivirus es incapaz de detectar el 100% de malware existente; por tanto, el poder disponer de múltiples herramientas de detección y análisis de virus mediante técnicas en la que se garantiza la no incompatibilidad, aumenta la calidad referida en términos de seguridad.

Como se ha tratado de inculcar durante todo el proyecto, existen diferentes vías gratuitas para la detección de ficheros maliciosos independientemente de su naturaleza (virus, troyanos, keyloggers, …) que ayudan a las organizaciones a disponer de un control total sobre el estado de salud de cada uno de los dispositivos que poseen. Son soluciones útiles que no suponen gasto alguno y permiten la personalización y adaptación a futuros desarrollos para cubrir nuevas necesidades.

Este tipo de métodos ‘externalizan’ el análisis fuera de la máquina a estudio. Ello permite que el dispositivo analizado no se vea afectado en materia de procesamiento de datos ni de memoria.

Junto con los objetivos logrados en el presente Trabajo Fin de Máster, se ofrecen líneas de innovación, investigación y desarrollo, como las que se indican a continuación:

- Detección de malware mediante varias vías, a través de:
  - Semántica
  - Comportamiento
  - Basado en técnicas heurísticas y bayesianas

- Patrones de clasificación

- Estudio del flujo de información de todo el sistema informático para la detección y análisis de malware
- Nuevas posibilidades de las firmas de virus para su detección y desinfección, así como la mejora en su eficiencia
- Detección de código malware ofuscado
- Aprovechamiento de fuentes de conocimiento públicas y gratuitas, alternativas a *VirusTotal*, como *Metascan* y *VirSCAN*
- Mejoras en los tiempos de respuesta de *VirusTotal* y/o estudios de método dinámico para determinar el tiempo exacto de cómputo
- Sistemas inteligentes de detección de malware
- Detección de malware en capas inferiores a la de aplicación (como capas física, red y transporte)
- Arquitecturas colaborativas para la detección y análisis de malware
- Técnicas para la disminución de falsos positivos y falsos negativos

Referencias bibliográficas

[1] AV-Comparatives. (2015). *Whole Product Dynamic "Real-World" Protection Test*. Recuperado de http://www.av-comparatives.org/wp-content/uploads/2015/07/avc_prot_2015a_en.pdf

[2] Kaspersky Lab. (2014). *Qué es un hash y cómo funciona*. Mensaje publicado en https://blog.kaspersky.com.mx/que-es-un-hash-y-como-funciona/2806/

[3] Granados Paredes, G. (2006). Introducción a la criptografía. *Revista Digital Universitaria, 7*(7), 1-17. Recuperado de http://www.revista.unam.mx/vol.7/num7/art55/jul_art55.pdf

[4] Krawczyk, H., IBM (s.f.). *Funciones unidireccionales y hash*. Material no publicado. Recuperado el 5 de agosto de 2015 de http://www.criptored.upm.es/intypedia/docs/es/video14/DiapositivasIntypedia014.pdf

[5] Google compra VirusTotal, una empresa española de antivirus http://www.abc.es/20120910/tecnologia/abci-google-compra-virustotal-espanola-201209101109.html

[6] Reventós, L., El País. (2012). *Google compra VirusTotal*. Recuperado el 10 de agosto de 2015 de http://tecnologia.elpais.com/tecnologia/2012/09/10/actualidad/1347276010_539192.html

[7] JSON. (s.f.). *Introducción a JSON*. Recuperado el 15 de agosto de 2015 de http://json.org/json-es.html

[8] McAfee, Inc. (2014). *Uso del archivo de prueba antimalware EICAR*. Recuperado el 16 de agosto de 2015 de https://kc.mcafee.com/corporate/index?page=content&id=KB59742&actp=null&viewlocale=es_ES

[9] Kaspersky Lab, (2013). *¿Qué es el virus de test EICAR? ¿Para qué está hecho?*. Recuperado el 16 de agosto de 2015 de http://support.kaspersky.com/sp/viruses/general/459

[10] Kaspersky Lab. (s.f.). *Seguridad 101: Los tipos de malware*. Recuperado el 15 de agosto de 2015 de http://support.kaspersky.com/sp/viruses/general/614

[11] VirusTotal. (s.f.). *Frequently Asked Questions*. Recuperado el 10 de agosto de 2015 de https://www.virustotal.com/es/faq/

[12] Movistar. (s.f.). *Amenazas a nivel de contenidos*. Recuperado el 16 de agosto de 2015 de http://www.movistar.es/particulares/internet/seguridad/terminos-habituales/

[13] Chicano, E. (2914). Gestión de incidentes en seguridad informática. IFCT0109. Recuperado de https://books.google.es/books?id=y63KCQAAQBAJ&lpg=PT63&ots=zmBDn6F9jK&dq=falsos%20negativos%20en%20seguridad%20informatica&hl=es&pg=PT63#v=onepage&q=falsos%20negativos%20en%20seguridad%20informatica&f=false

Anexo A – Código fuente y licenciamiento

Un repositorio es un depósito centralizado donde se almacena información digital.

Para la elaboración de este Trabajo Fin de Máster, se han creado tres repositorios públicos en la plataforma de desarrollo colaborativo *GitHub*[vii].

GitHub, además, permite disponer de un control de versiones. Personas o entidades que deseen efectuar modificaciones en el código fuente podrán hacerlo registrando los cambios realizados sobre un archivo o conjunto de ellos a lo largo del tiempo.

Los tres repositorios desarrollados para este trabajo han sido publicados con licencia *GNU General Public License versión 2* y se encuentran bajo la dirección https://github.com/dvdrodriguez.

A continuación se detallan cada uno de ellos:

- **Perl**
  - *URL*: https://github.com/dvdrodriguez/Perl
  - *Descripción*: Código fuente completo del detector del malware. Incluye las librerías "*VirusTotal.pm"* y "*VTMySQL.pm"*, así como el programa principal, "*search.pl"*.
- **MySQL**
  - *URL*: https://github.com/dvdrodriguez/MySQL
  - *Descripción*: Código fuente SQL para la creación de la base de datos y tablas asociadas. Se trata de un elemento fundamental para el correcto funcionamiento del software implementado.
- **PHP**
  - *URL*: https://github.com/dvdrodriguez/PHP
  - *Descripción*: Código fuente desarrollado en PHP para la realización de la parte optativa (es decir, no esencial para la ejecución de la solución propuesta). En ella, se muestran los resultados obtenidos fruto de los análisis que haya realizado el detector, pudiendo ser efectuados dispositivos multiplataforma y ubicación descentralizada.

[vii] https://github.com/

De igual forma proporciona -a modo de ejemplo- múltiples gráficas que evalúan la actividad de los antivirus que han participado en dichos análisis.

Observación: La licencia *GNU GPL v2* aplica, en exclusividad, al código fuente desarrollado para el presente trabajo y, por tanto, no exime de responsabilidades, modifica o anula las condiciones de uso del servicio *VirusTotal* que se encuentra debidamente descrito en https://www.virustotal.com/es/about/terms-of-service/.