

# Evaluando la Evaluación

Evaluación de programas antimalware para la empresa

Autores: Andrew Lee, Chief Research Officer de ESET y David Harley,  
Research Author de ESET

Fecha: Lunes 25 de Agosto del 2008

## Sobre los Autores

### David Harley

Profesional certificado en seguridad de sistemas (CISSP), autor e investigador de ESET, es un respetado investigador de programas antivirus con una amplia experiencia y posee títulos en asesoramiento sobre seguridad, gestión de servicios con ITIL e informática médica. Hasta el año 2006 trabajó en el Servicio de Salud Nacional del Reino Unido, donde se especializó en el manejo de programas maliciosos y todas las formas de abuso por medio de correo electrónico; también dirigió el Centro de Asesoramiento de Amenazas. Ha trabajado como escritor independiente y consultor en industrias antivirus y de seguridad, y es Chief Operating Officer de la Red de Intercambio de Información Antivirus (AVIEN, por sus siglas en inglés).

Fue coautor del libro "Viruses Revealed" ("Los virus: desenmascarados") y colaboró con capítulos de muchos otros libros sobre seguridad y educación para las editoriales más importantes, además de haber redactado una gran cantidad de artículos y discursos para conferencias. Es el editor técnico y autor principal del libro "The AVIEN Malware Defense Guide for the Enterprise", ("La guía AVIEN para la defensa contra códigos maliciosos para la empresa"), editado por Syngress.

### Andrew Lee

Profesional certificado en seguridad de sistemas (CISSP), es Chief Research Officer de ESET LLC. Fue uno de los miembros fundadores de la Red de Intercambio de Información Antivirus (AVIEN) y su organización hermana: Información Antivirus y Sistema de Alerta Temprana (AVIEWS); es miembro de la Asociación de Investigadores Antivirus de Asia (AVAR) y periodista de la organización WildList. Previamente trabajó como administrador de sistemas en una importante entidad gubernamental británica para la defensa contra códigos maliciosos.

Andrew fue uno de los contribuyentes principales de la Guía AVIEN, y es autor de numerosos artículos sobre códigos maliciosos. Es orador frecuente en conferencias y eventos, entre los que se incluyen los seminarios de ISC2, AVAR, Virus Bulletin y EICAR.

*Este artículo fue escrito para ser presentado en la 10ª Conferencia Anual Internacional de la Asociación de Investigadores Antivirus de Asia (AVAR, por sus siglas en inglés) en Seúl en el año 2007.*

## Contenidos

- **Sobre los autores**
- **Resumen**
- **Introducción**
- **Probando nuestra paciencia**
- **Leer entre líneas los análisis comparativos**
- **Equipos y tácticas**
- **Empresas antimalware contra el mundo**
- **Empresas antimalware y las demás industrias de seguridad**
- **La ética en la evaluación de programas antivirus**
- **Aspectos técnicos**
- **Verificación de muestras**
- **Colecciones de muestras**
- **Detectar la falacia**
- **¿Cuán práctica es la evaluación “hágalo usted mismo”?**
- **Los virus no son todo el problema**
- **¿Qué se necesita para hacer una evaluación satisfactoria?**
- **¿Quiénes son los evaluadores confiables?**
- **Configuración predeterminada**
- **Conclusión**
- **Referencias**
- **Recursos adicionales**

## Resumen

Los programas antimalware siguen siendo un elemento de defensa esencial para la mayoría de las empresas, que comprensiblemente ansían obtener el equilibrio adecuado entre costo y efectividad. Lamentablemente, los periodistas, grupos de consumidores y aficionados en temas de seguridad cada vez encuentran maneras más creativas e inapropiadas de evaluar los programas especializados en la detección. En este artículo, trataremos la siguiente serie de temas principales:

1. Leer entre líneas los análisis comparativos
2. Programas antivirus y antimalware contra el mundo
  - La ética al evaluar productos
  - Confiabilidad y competencia
  - Realidad y ficción en la opinión pública sobre la industria antimalware
  - Lo que el resto de la industria de seguridad no comprende
3. Aspectos técnicos de las evaluaciones:
  - Verificación de muestras
  - Evaluaciones con códigos maliciosos replicativos
  - Evaluación proactiva (retrospectiva) y heurística
  - Evaluación de la actualización del producto
  - Evaluación de códigos maliciosos activos en el mundo real (in-the-Wild)
  - Códigos maliciosos no replicativos
  - Evaluación en tiempo real versus evaluación bajo demanda
  - Evaluación de falsos positivos
4. Evaluando a los evaluadores: recursos satisfactorios versus recursos insatisfactorios
  - Evaluación y validación
  - Revisores especializados
  - Evaluación tercerizada
  - Las críticas persistentes del aficionado en temas de seguridad y del experto instantáneo
5. Las ventajas y desventajas de la evaluación "hágalo usted mismo": ¿cuán práctica es?

## Introducción

Como ocurre con la nomenclatura para designar códigos maliciosos y la supuesta dependencia total de los programas antivirus en las firmas de virus para efectuar la detección, la evaluación del rendimiento de detección de los programas antimalware, en especial la evaluación comparativa, es constantemente una fuente de gran controversia. También es un tema importante que se trata en los artículos escritos, debido a una reciente evaluación comparativa controversial llevada a cabo por Untangled.com en LinuxWorld [1; 2]. Con razón, este aspecto de la evaluación de productos se considera de gran importancia:

- Las medidas para controlar la invasión y el impacto de los códigos maliciosos, ya sea en forma de un programa antivirus convencional o como parte de un sistema general para prevenir intrusiones, son un elemento de defensa esencial dentro de la empresa.
- No “son todos iguales”: a pesar de que el espectro de virus conocido detectado por la corriente dominante de analizadores antivirus es consistente, la diversidad de las demás funciones y de otros tipos de detección de códigos maliciosos varía ampliamente. Esto ocurre en particular en la detección de amenazas nuevas (nunca antes vistas).
- La mayoría de las empresas e individuos necesitan conseguir un equilibrio entre costo y efectividad.

## Probando nuestra paciencia

Lamentablemente, algunos periodistas, grupos de consumidores y aficionados en temas de seguridad cada vez encuentran maneras más creativas e inapropiadas para evaluar los programas especializados en la detección. Algunos de los problemas centrales que hemos notado en análisis comparativos en el transcurso de los años incluyen:

- Paquetes para evaluaciones con supuestos virus que en realidad no lo son y malware no viables, como virus defectuosos, archivos basura, archivos inocuos y archivos de prueba inofensivos.
- Códigos maliciosos simulados. Pueden traer diversas complicaciones, como la paradoja básica de que un analizador sea “recompensado” por diagnosticar incorrectamente que dicha simulación, es en efecto, el código malicioso al que pretende imitar (recuerde que el propósito de un detector de códigos maliciosos es detectar únicamente códigos maliciosos).
- Paquetes de códigos maliciosos que, con frecuencia, generan muestras no viables [2, 3].
- Códigos no maliciosos o códigos maliciosos inapropiados dentro del contexto particular: por ejemplo, evaluar analizadores web con muestras de HTML que sólo aparecieron activos en el mundo real como transmisiones SMTP [4], o el uso del archivo de prueba EICAR incrustado incorrectamente en un documento de Word [5].
- Muestras no validadas que se toman por códigos maliciosos (en general porque han sido identificadas como una amenaza específica, como una detección genérica reconocible o como código “sospechoso”) [6].
- Validación “circular” (el código malicioso se “valida” probándolo con uno de los productos en evaluación) [6, 7].
- ¿Manzanas o naranjas?: pruebas comparativas donde productos cuyas funciones, niveles de configuración y muchas otras características difieren significativamente se evalúan usando el mismo paquete de prueba y una metodología básica. Por ejemplo, evaluar analizadores bajo distintos sistemas operativos o sin tener en cuenta si han sido diseñados para estaciones de trabajo, servidores con redes de área local (LAN) o ubicaciones perimetrales, o la diversidad de servicios protegidos [1, 6].
- Objetivos de evaluación indefinidos: por ejemplo, no establecer una clara diferencia entre pruebas de detección heurística, filtrado genérico, identificaciones casi exactas, etc. [6, 7, 8].

Nos han comentado que no es necesario ser cocinero para saber si una comida sabe bien. Sin embargo, nosotros rebatimos que sí es necesario entender algo de nutrición para saber si una comida es en realidad saludable...

## Leer entre líneas los análisis comparativos

Las evaluaciones deficientes raramente se cuestionan con excepción del contexto de la industria antimalware, a la que se le adjudican razones siniestras (incluso por los mismos evaluadores) para no querer que se realice ninguna prueba o, al menos, ninguna que ella no pueda de algún modo controlar. Las razones para haber llegado a esta conclusión no son inexistentes: las empresas antivirus suelen adquirir las colecciones de muestras más completas, hacer validaciones de rutina y clasificar los programas sospechosos como parte del proceso interno; y no comparten los resultados con facilidad. Al menos, algunos de los motivos para mostrar esta reticencia son por completo respetables; no obstante, es demasiado fácil para el público (o los “líderes del pensamiento” que influyen en la percepción pública) interpretarlo como medida autoprotectora y de interés propio [9].

En este punto surge una curiosa dualidad: con frecuencia se asume que las empresas de este sector de la industria pretenden obtener una ventaja competitiva manteniendo las muestras en secreto, creando sus propios códigos maliciosos, etc., incluso algunos creen que cierran filas y conspiran entre ellos por el bien de la industria y en detrimento del bien común [10]. No podemos asegurar que dos (o más) individuos de empresas antivirus no hayan conspirado alguna vez en formas siniestras, monopolizadoras, secretas; pero en rara ocasión tenemos razones para sospechar que existe una conspiración de tal índole. No obstante, sí nos resulta extraño lo difícil que es convencer a las personas ajenas a la industria del gran espíritu de cooperación que existe entre los investigadores a través de los perímetros corporativos, en particular cuando se trata de compartir muestras entre individuos de confianza.

“...más allá de sus avances técnicos, la industria antivirus aún no llega a ganar los corazones y las mentes del público. Por el contrario, nuestros clientes, los medios de comunicación y en especial los demás sectores de la industria de seguridad, desconfían de nosotros. Según parece, nos califican de incompetentes, elitistas, conspiradores, interesados en el dinero, codiciosos por obtener publicidad y, en general, desprovistos de ética. Pero nosotros también tenemos el derecho a tener nuestros defectos.” [9]

Para el individuo informado, los distintos focos de problemas que se mencionan en la sección anterior indicarán qué programa es incompetente con la misma certeza que los actores de la serie televisiva “CSI” identificarían a un asesino en la vida real. No obstante, cuando los informes de evaluaciones se basan en este tipo de supuestos y estereotipos, este prejuicio además sugiere una competencia cuestionable y una falta de conocimiento general sobre el tema. Sin duda, existen organizaciones evaluadoras muy bien reconocidas y aceptadas que objetarán este argumento. Dichas organizaciones y evaluadores se las arreglan para mantenerse independientes de la industria, a pesar de encontrarse ampliamente sumergida en ella. Por más sorprendente que parezca, la mayoría de los investigadores de antivirus prefieren ver buenas evaluaciones por diversos motivos.

En primer lugar, un buen producto se destacará en una buena evaluación y es probable que no le vaya demasiado mal en una evaluación deficiente. Este hecho genera una gran frustración en los vendedores, ya que saben que la calidad de su producto no está siendo correctamente reflejada en la evaluación.

En segundo lugar, una buena evaluación es capaz de revelar de manera legítima fallas y puntos débiles en los productos antimalware, para conveniencia de vendedores (y clientes), que así tendrán oportunidad de rectificarlo.

En tercer lugar, obtener buenos resultados en una evaluación respetable es una excelente publicidad. Una vez más, muchos se sorprenderán al saber que algunos vendedores no usarán los resultados de ciertas evaluaciones para marketing (por más que el resultado haya sido "bueno") porque la reputación y calidad del producto no mejorarían al obtener un nivel elevado en una evaluación deficiente (y seguramente será utilizado contra el vendedor por sus competidores).

En cuarto lugar, los buenos evaluadores y las buenas evaluaciones hacen que los vendedores se mantengan honestos. Si los vendedores sólo usaran los resultados de evaluaciones deficientes, prácticamente no sería necesario que hicieran un producto decente, sólo deberían ir retocando su producto para que logren pasar las pruebas. Una evaluación real de las capacidades de un producto es beneficiosa para todos. Por el contrario, una evaluación deficiente perjudica a los clientes del vendedor que la realizó, y también a la comunidad más amplia de usuarios de programas antimalware, ya que malinterpreta (en forma favorable o desfavorable) las capacidades reales de los productos. Los productos antimalware usan algunas de las tecnologías más avanzadas y complejas entre los programas informáticos modernos y requieren un esfuerzo que dista de ser pequeño para analizarlos y evaluarlos de manera correcta.

Quizá sea posible llevar a cabo una evaluación útil sin ser un experto mundial en programas antivirus, pero seguramente no sea posible hacerlo sin tener una idea razonable de cómo funcionan las tecnologías de códigos maliciosos y los programas que los detectan. Además, las dudosas teorías conspirativas, más allá de lo adecuadas que resulten para tramas de películas, conforman una base aún más pobre para una evaluación rigurosa que los principios científicos.

Sin embargo, también hay otros indicadores de un desempeño deficiente.

## Equipos y tácticas

Los terceros que distribuyen servicios antimalware (servicios tercerizados, motores renovados, productos con motores múltiples) pueden considerarse parte de la industria antimalware, pero su conocimiento tecnológico de las amenazas y las formas de detectarlas en general es sorprendentemente básico [11]. Con mucha frecuencia, el componente antimalware del servicio es, en esencia, una caja negra con un motor no identificado en un paquete de una marca específica. En casos extremos, ni el distribuidor del servicio ni el cliente son capaces de personalizar o configurar en forma significativa las funciones básicas, ya sea porque carecen de los conocimientos necesarios o porque la empresa que creó la aplicación no la diseñó para darles el acceso correspondiente. Por lo tanto, cuando un tercero publica los resultados de su propia evaluación, sería demasiado ingenuo confiar en su competencia y, más aún, en su imparcialidad.

De todas formas, no cabe duda de que el patrocinio y la evaluación del vendedor pueden sugerir un conflicto de intereses. En un caso reciente [1], un proveedor de servicios que incluía el filtrado de códigos maliciosos realizó unas pruebas con una serie de productos entre los que se encontraba el que ya estaba incorporando a su servicio. Aunque no sugerimos una mala práctica deliberada, existe un riesgo, en ese caso, de que el evaluador sobrevalúe la capacidad de su producto [12] y se incline a favor del programa que ya está utilizando (en especial cuando dicho programa resulta ser gratuito): después de todo, los resultados que indican que el producto en cuestión no cumple con los mismos estándares que otros productos o no los excede tendrán un impacto negativo para el marketing de la marca que lleva el producto. Está claro que los creadores de productos antimalware comerciales (de núcleos), por lo general son perfectamente capaces de realizar evaluaciones comparativas competentes y casi siempre lo hacen de manera regular para corroborar el rendimiento de su producto comparado con los de la competencia. No obstante, habitualmente evitan los dilemas de incertidumbre ética y los riesgos de generar un marketing negativo ocultando los resultados y manteniendo una distancia discreta de las organizaciones evaluadoras de buena reputación, aún cuando cooperan con ellas.

Las metodologías de evaluación y validación inapropiadas o indefinidas constituyen una señal de advertencia fundamental. A pesar de que en general es poco práctico hacer un detalle demasiado exhaustivo de cada evaluación, las afirmaciones del tipo: “tomamos virus de sitios web de crackers y los ejecutamos con el producto en evaluación” o el silencio total referente a cómo se realizó la evaluación deben considerarse con una desconfianza extrema [13].

Los paquetes de muestras para evaluación muy pequeños, pocas veces tienen lugar en las evaluaciones competentes, aún cuando las muestras involucradas hayan sido validadas correctamente. Puede haber excepciones [14], pero en ese caso la responsabilidad recae en el evaluador, que deberá dejar bien en claro cuáles son las limitaciones de su enfoque y por qué es apropiado en ese caso en particular. Un punto

de vista indica que “si se hace una evaluación con cien virus de los cuales sólo tres están activos en el mundo real, son esos tres en los que estoy interesado”. Este punto de vista no es del todo inválido, pero pasa por alto cuestiones importantes:

- Al aceptar este punto de vista, uno debe estar seguro de que lo que está evaluando está realmente activo en el mundo real (In-the-Wild). No es en absoluto tan simple como le parece al evaluador novato.
- Los productos antivirus y antimalware no deben detectar sólo lo que está activo en el mundo real: deben detectar lo que en el pasado solía estar activo en el mundo real (porque puede emerger en un sistema obsoleto, en dispositivos de memoria que se vuelven a utilizar después de mucho tiempo, etc.) y también deben detectar códigos maliciosos que se sabe que existen en laboratorios pero nunca han estado activos en el mundo real (como las colecciones zoo) por las dudas de que de pronto “tengan suerte” y se liberen en el mundo real.

¿Qué es lo que se evalúa en realidad? Hay un tipo de análisis comparativo equivocado al que se lo suele conocer como la evaluación de “naranjas y manzanas” (o viceversa), porque implica tratar objetos muy diferentes como si fueran idénticos en cuanto a forma y función; en consiguiente se asume que la misma metodología sirve para evaluarlos. No entraremos en detalles sobre los diferentes tipos de evaluaciones de rendimiento para no repetir los temas que se tratarán en otras de las presentaciones de este evento [15]. (Por razones similares, no pretendemos dar una definición exhaustiva de estrategias de evaluación y soluciones [16] y les recomendamos los demás artículos relacionados con el tema). Sin embargo, no podemos dejar de lado este tema sin haber resaltado la necesidad de comprender las diferencias entre distintos tipos de evaluaciones de rendimiento y los peligros de confundirlos sin el conocimiento adecuado. Para dar un ejemplo que ya fue mencionado recientemente [1], se tomará la evaluación que intentó alcanzar varios blancos usando la misma flecha [6]:

- Incluía explorador de dispositivos, explorador de puerta de enlace, explorador de correo, explorador de la estación de trabajo, sin tener en cuenta la plataforma ni la interfaz (por interfaz gráfica de usuario o por línea de comandos), todo en la misma evaluación.
- También combinaba (sabiéndolo o no) muchos tipos de evaluaciones en un único análisis:
  - Reconocimiento del archivo de evaluación EICAR (al parecer basado en la incorrecta suposición de que el hecho de reconocer EICAR.COM prueba que la configuración es correcta) [1, 6, 13].
  - Reconocimiento de supuestos códigos maliciosos activos en el mundo real. Como es poco probable que el evaluador haya tenido acceso a las muestras de la WildList (<http://www.wildlist.org>), suponemos que los códigos maliciosos fueron “validados” al ser identificados por uno o más programas antimalware como los códigos maliciosos que aparecen en la lista WildList o una fuente similar – por supuesto, esto no cumple con un estándar aceptable de validación, identificación o recopilación [17; 18].

- Reconocimiento de supuestos códigos maliciosos que no se cree que estén activos en el mundo real, pero que “necesariamente” deben ser identificados por los analizadores en evaluación (evaluación de colecciones zoo).
- Reconocimiento de supuestos códigos maliciosos que no se espera que sean conocidos para el analizador (en este caso, amenazas zero day y códigos maliciosos “creados para la ocasión” aportados por la audiencia) En este caso, parece que las muestras directamente no fueron validadas como maliciosas o replicativas y el evaluador admitió que, de hecho, él no sabía lo que eran. Por zero day bien puede haber querido decir muestras de códigos maliciosos demasiado recientes para poder ser identificados como variantes específicas. Suponemos que las muestras creadas especialmente, son muestras existentes alteradas o cuyo código se volvió a escribir. Podríamos catalogarlo como un intento de evaluar la heurística – es decir, la efectividad de un explorador para detectar códigos maliciosos aún no identificados por medio del reconocimiento de características que indican un comportamiento malicioso. No obstante, la completa ausencia de validación en este caso significa que lo que hace, en realidad, es evaluar la habilidad de un explorador de detectar lo mismo que otros analizadores: de esta forma, el analizador bajo evaluación “triumfa” si identifica objetos como maliciosos (o como sospechosos) que al menos un explorador más también va a detectar, más allá de que sea malicioso o viral de verdad. Lamentablemente, este es un ejemplo extremo de una metodología equivocada de evaluación, que trae a memoria la investigación parapsicológica de Rhine en la universidad de Duke [19]; en lugar de la evaluación de programas antimalware en Hamburgo [20] o Magdeburgo [21].

Lamentablemente, existe cierta evidencia de que estas evaluaciones deficientes (para no mencionar la cantidad, cada vez mayor, de muestras nuevas que deja de ser manejable) ha originado el fenómeno al que denominamos “detección por copia en cascada”, donde un objeto determinado es detectado (ya sea como falsa alarma o no) por un explorador antimalware y, en consecuencia, es agregado a las listas de detecciones de muchos otros exploradores. En muchos casos, éste parece ser un proceso en gran medida automático, donde algunos vendedores simplemente usan los exploradores de otros vendedores para determinar si un código es malicioso – un método de categorización por el cual, a su vez, castigan a los malos evaluadores. Este es un debate un tanto tangencial que volveremos a tratar en otro artículo, donde investigaremos este fenómeno con mayor profundidad. Deberán contentarse con saber que este tipo de comportamiento ciertamente no es útil y, por el contrario, ha llevado a una replicación vergonzosa de falsos positivos a través de los analizadores antimalware.

## Empresas antimalware contra el mundo

Desde hace tiempo, hemos estamos fascinados con el fenómeno de la ambivalencia pública hacia la industria antivirus/antimalware [9]. Por un lado, existe la creencia común de que prácticamente cualquier persona sabe más o es más confiable en lo que respecta a los códigos maliciosos y a su manejo que la industria antimalware, incluyendo hackers (en el sentido peyorativo) y creadores de códigos maliciosos, aficionados en temas de seguridad y buscadores de vulnerabilidades, y prácticamente cualquier profesional en seguridad ajeno a las industrias antimalware. Por otra parte, se da por sentado que ejercen una influencia siniestra y de interés propio sobre entidades evaluadoras y otros grupos que se espera que sean imparciales.

Existe una larga serie de mitos populares y afortunadamente, ideas equivocadas diseminadas por todos esos grupos, sobre los productos y las personas que los crean y los mantienen:

- Que sólo les preocupa detectar virus, y no códigos maliciosos en general. (Les advierto, conocemos el caso de proveedores que intentan liberarse de las cláusulas punitivas asegurando que el código malicioso que no detectaron era un gusano y no un virus y que, por lo tanto, no estaban obligados bajo contrato a detectarlo [22].)
- Que no son confiables porque ellos son los que crean todos los virus (¡todavía se sigue creyendo!).
- Que los vendedores son ambiciosos porque insisten en cobrar por sus productos cuando todos “saben” (y las evaluaciones deficientes “prueban” o sus resultados se malinterpretan como una prueba de) que los antivirus gratuitos son mejores (¡!).
- Que la industria aún tiene la reputación de estar obsesionada con la detección por firmas y la protección de esta fuente de ingresos, a pesar de los avances considerables en la tecnología heurística desde la década de 1990.
- O que son simplemente incompetentes, ya que no son capaces de detener todos los códigos maliciosos (y de paso traer la paz mundial en sus ratos libres).

Con el transcurso de los años, una cantidad considerable de instancias de “lo que todos saben” ha estado circulando en lo que respecta específicamente a la evaluación:

- Que el “establecimiento” de evaluación es un esclavo y es en esencia inseparable del establecimiento comercial.
- Que los evaluadores establecidos dependen, para sus ganancias, de los valores entregados por el establecimiento comercial y que esto genera una preferencia en detrimento de los vendedores pequeños, los vendedores de códigos abiertos, etc. Por ejemplo: “Debo asumir que los laboratorios de evaluación no son imparciales al realizar las evaluaciones, probablemente porque obtienen sus ingresos de los vendedores comerciales que les pagan para evaluar. Sin duda, sus clientes no estarán satisfechos si los laboratorios de evaluación aseguraran que una solución de código abierto y gratuito fuera mejor” [1].
- Que son imprecisos sobre su metodología intencionalmente.
- Que se concentran en evaluaciones que perpetúan enfoques irreales u obsoletos para favorecer el interés de la industria en lugar de beneficiar al cliente.

Es claro que hay más verdad en algunos de estos mitos e ideas erróneas que en otros. Por ejemplo, algunas de las quejas sobre metodologías ineficaces usadas en algunos establecimientos se pueden relacionar con la desconfianza de los primeros protocolos para evaluación del Centro Estadounidense de Aplicaciones de Supercomputación (NCSA, según sus siglas en inglés) [24; 25]. Otras quejas se relacionan directamente con la inquietud respecto a si las muestras de la WildList/WildCore, al menos en su forma actual, son apropiadas como recurso básico al evaluar la detección [26; 27; 28].

Ahora, como es un punto conveniente para un apartado de este tipo, demos cierre al mito, que desde la lógica es una falacia, pero que igual es irritante por su persistencia, de que las empresas crean los códigos maliciosos. Esta es la pregunta que más nos hacen miembros de la audiencia cuando se enteran cuales nuestra área de especialización. Además de la sonrisa cansada y la rápida negación, con frecuencia seguidas de un suspiro y una respuesta sobre lo mucho que nos gustaría tener más tiempo para tomar sol en las playas en lugar de estar enterrados hasta la cabeza analizando códigos maliciosos, existe otra réplica más sensata y evidente. Por un lado, el público espera que los programas antivirus detecten el 100% de todos los códigos maliciosos, todo el tiempo. Sin embargo, según la experiencia del público y como fue comprobado en gran cantidad de evaluaciones de distintas fuentes, no es lo que ocurre. De hecho, una queja típica que llega a los departamentos de soporte para vendedores es: “su producto no detectó el virus xxx en mi sistema”. Sin duda alguna, si existiera un “departamento de creación de virus” tendría la obligación de proveer a los vendedores la detección de dicho código malicioso mucho antes de liberarlo, para que el cliente experimente el beneficio que le brinda su protección integral. Por supuesto, no sólo es totalmente descabellado sugerir que los vendedores fabrican el código malicioso, también es obvio que si lo hicieran estarían cometiendo un suicidio comercial. Aún así, este tipo de teorías

conspirativas persisten y, a pesar de la gran cantidad de gastos y de logística que se necesitarían para llevar a cabo dichas tareas (aún mayor en escala que lo que habría necesitado la NASA para recrear la llegada de tres hombres a la Luna en 1969), es poco probable que desaparezcan pronto. A pesar de todo, una cosa es cierta: los evaluadores de programas antivirus han creado públicamente más virus “nuevos” que los que alguna vez pueden haber sido creados por la comunidad de empresas antimalware.

Lamentablemente, esta situación tiene todas las de perder: con frecuencia nos sugieren que el hecho de no probar sus propios productos con los nuevos malware que crearon es un síntoma de la incompetencia de la industria antimalware. ¿Cómo sabemos que no es así? En realidad, estamos bastante seguros de que algunos investigadores de la industria generan ataques de códigos maliciosos para realizar “pruebas de concepto”, pero bajo condiciones estrictamente controladas por personas sumamente concientes de los riesgos éticos así como prácticos. Lo que no hacen es publicar los resultados de aquellas evaluaciones como ejercicio publicitario porque sería claramente un engaño. Por supuesto, es prácticamente imposible convencer a algunas personas de que la industria antivirus no controla en forma directa la industria de la evaluación.

## Empresas antimalware y las demás industrias de seguridad

Otro de los problemas es que la industria antimalware tiene una reputación bastante mala frente al resto de las industrias de seguridad, que en forma consistente no logran entender los siguientes puntos [2]:

- Los códigos maliciosos no son una especialidad sencilla y quienes trabajan en otras especialidades no cuentan necesariamente con una comprensión profunda e instintiva sobre la tecnología de códigos maliciosos, los programas para detectarlos y los problemas relacionados a su gestión y cultura.
- Sus suposiciones de veinte años de antigüedad respecto al hecho de que la tecnología de detección se basa íntegramente en firmas no tienen fundamentos; lo que una vez más se basa en una comprensión deficiente de las realidades de tecnologías modernas de códigos maliciosos y programas antimalware.
- La costumbre de dar opinión sin el conocimiento necesario y el síndrome de la falsa autoridad [29] viven y se sienten a gusto en el instituto SANS, además de otros lugares, donde Alan Paller elogió una evaluación deficiente implementada por Consumer Reports por haber “ayudado mucho a mejorar la investigación del producto” al probar que “los vendedores de antivirus no detectan ni bloquean los virus con rapidez” [30; 2].
- El choque de culturas entre el modelo de transparencia completa de metodologías usado por la mayor parte de la industria de seguridad y el modelo históricamente reservado de la industria antivirus [31] sigue afectando la relación entre especialistas en programas antimalware y otros sectores de la industria, sin mencionar los demás grupos influenciados por esos sectores (la prensa, el público, los que pretenden saber del tema y dan su opinión).

## La ética en la evaluación de programas antivirus

La industria antimalware con frecuencia se queja arduamente de las evaluaciones deficientes, pero no realiza un buen trabajo a la hora de explicar cuáles son sus objeciones. Por fuera de este sector específico de la industria, pocas personas comprenden las objeciones éticas que plantea la industria sobre la creación de programas replicativos con el propósito de hacer evaluaciones [32]: esto se traduce como “Ellos dicen que no es ético evaluar los códigos creados porque no quieren que nos demos cuenta de que sus programas son basura”. Por supuesto, dejemos en claro cuáles son las cuestiones éticas reales, pero quizás haya puntos en los que haga falta hacer aún más hincapié:

- A pesar de que las objeciones éticas y de seguridad (por más que no sean de ninguna manera triviales) no logren convencer a la corriente predominante de seguridad [33, 34] o a los que pretenden saber de seguridad y responden a las entradas sobre antivirus que aparecen en blogs [35], nuestra experiencia indica que a veces es más fácil convencer desde un nivel técnico. Después de todo, mientras que no todos, incluso en la industria antimalware, creen que jamás es justificable crear códigos maliciosos replicativos con el único propósito de hacer evaluaciones o para investigación, incluso bajo condiciones controladas, sería más difícil argumentar que no existen dificultades morales o éticas al engañar a la prensa o al público [36], ya sea en forma deliberada o no intencional, usando metodologías inapropiadas o creadas deficientemente.

La industria antimalware no podrá ganar el corazón y la mente de las personas al fomentar la impresión de que todos los intentos de evaluar la detección antimalware se rechacen en forma desconsiderada. Para alguien que se halla fuera del círculo privilegiado de algunos investigadores independientes confiables, siempre fue casi imposible evaluar ciertas características de productos antimalware en un nivel que la industria considere aceptable. En líneas generales, los motivos históricos son honestos, pero al mundo le resulta extraño que la industria use este “elitismo” para invalidar cada evaluación que arroje resultados inesperados.

## Aspectos técnicos

Prosigamos ahora con una visión general de algunos aspectos más técnicos de la evaluación. No nos adherimos a la idea de que sólo una elite de profesionales es capaz de decir algo útil sobre el rendimiento de los programas antivirus. Pero sí afirmamos que uno no hace una evaluación aceptable desde las ideas erróneas, las deducciones ilógicas o el síndrome de la falsa autoridad. Quien nada sabe sobre técnicas de evaluación o códigos maliciosos, tendrá muy baja probabilidad de efectuar una evaluación válida. Incluso quien cuente con los conocimientos no los podrá aplicar correctamente sin el tiempo y los recursos necesarios.

Hace poco, a uno de los escritores le llamaron la atención en privado por criticar la metodología de la evaluación de Untangled [1, 6] sin haber probado el paquete de muestras por cuenta propia. Este paquete de muestras estaba disponible en forma gratuita en el sitio web de Untangled, lo que en cierta forma fue problemático. De hecho, la situación se generó por un malentendido: el paquete de muestras fue examinado con suficiente detalle para identificar algunas muestras incorrectas: por ejemplo, archivos de 0 bytes. En su defensa, el escritor en cuestión argumentó lo siguiente:

- Uno de los principios de una buena evaluación es saber con qué estamos evaluando – es decir, es necesario validar las muestras. Sin embargo, nadie le estaba pagando a él para hacer la validación del evaluador: la validación, cuando se hace apropiadamente, es una actividad muy intensiva en cuanto al consumo de tiempos y recursos y, al verlo en retrospectiva, el escritor no tuvo ningún incentivo ni propósito útil para un gasto tan grande.
- Reproducir una evaluación defectuosa no tiene propósito excepto el de verificar que los resultados eran los que se habían informado: no valida la metodología por la cual se obtuvieron.
- En esta instancia, incluso si todas las muestras hubieran sido descartadas, no habría tenido ningún efecto material en las diversas fallas metodológicas presentes, como un paquete de muestras pequeño, la preferencia hacia un explorador incluido, así como configuraciones, elección de plataforma y objetivos de análisis inconsistentes.

El punto importante, es que la evaluación sensata tiene mucho que ver con conocer qué tipos de evaluaciones son posibles y útiles, con los recursos disponibles. Y por supuesto, saber que no hay que probar las muestras con recursos inapropiados como VirusTotal (que nunca se pensó con esa intención).

## Verificación de muestras

En la industria, la validación de muestras siempre [37] fue considerada un factor crítico en la evaluación apropiada de programas de detección [38] y, técnicamente, es muy demandante. La industria se aferra rápido a la creencia de que no se puede simplemente ejecutar uno o más exploradores de malware con una muestra y, si es identificada como un malware en particular, aceptar dicha “detección” como validación; en especial, si el programa antimalware usado con ese propósito es uno de los exploradores en evaluación.

Queda claro que, al proceder de esta forma, se introduce una gran parcialidad en la evaluación, ya que se confía en la competencia del proveedor del software y se asume que es más “correcto” que los antimalware que no concuerdan con él, sin tener en cuenta el riesgo de generar falsos positivos (o de detección de archivos dañados que tengan por casualidad las suficientes partes intactas para ser detectadas – algunos vendedores detectan esos archivos en forma deliberada para reducir la cantidad de correo basura que el cliente encontrará, por ejemplo, en la puerta de enlace del correo electrónico). Este enfoque tiene ventajas obvias cuando se realiza un ejercicio de marketing, pero no es aceptable como evaluación genuina e imparcial y demuestra la importancia de separar a la entidad evaluadora del vendedor o de cualquier otra persona con un interés inmanente en uno de los productos evaluados [39].

Como expresa Joe Wells [40]: “... una cuestión crítica que muchas veces se pasa por alto es la tendencia a sospechar de inmediato del producto antivirus cuando una muestra de virus no es identificada. Si observamos la calidad histórica de virus y de productos antivirus, sería más sensato que el evaluador sospeche de inmediato de la muestra de virus y no del producto antivirus. Es mucho más probable que la defectuosa sea la muestra, no el programa antivirus”. Sin embargo, esta referencia a problemas con un posible falso negativo no significa que cuando un solo explorador detecta una amenaza hay que sospechar que se trata de un falso positivo. Simplemente destaca la necesidad de que el evaluador sea escrupuloso respecto a (1) la calidad de la muestra – debe ser un programa genuinamente malicioso y/o replicativo, dependiendo del tipo de evaluación (2) la procedencia de la muestra – debe ser correctamente identificada como un ítem específico de malware y en su contexto correcto (por ejemplo, una sola variante/subvariante de la cual se desconoce si está activa en el mundo real, no debería considerarse o usarse como si fuera una muestra validada, originada por WildCore, que cumple con los criterios técnicos de los códigos maliciosos activos en el mundo real [41]).

## Colecciones de muestras

La verdadera validación requiere, entre otras cosas, que uno pruebe que está trabajando con una muestra replicativa viable (asumiendo que hablamos de virus, por supuesto – otros tipo de códigos maliciosos presentan otros problemas...) y que haya sido correctamente identificada como un

programa/variante/subvariante malicioso. Realizar esta tarea en forma correcta es difícil y lleva mucho tiempo, probablemente siendo una razón suficiente por la cual los aficionados casi nunca lo hacen. También es el motivo por el que las evaluaciones comparativas bien fundamentadas son costosas para montar y, por ende, no están disponibles para quienes no se subscribieron. Es también por eso que las evaluaciones que usan la lista WildList [6] todavía se llevan a cabo [7], a pesar de que las entradas de una lista específica representen sólo una escasa proporción de todos los códigos maliciosos conocidos [41], e incluso de códigos maliciosos que se sabe que han estado activos en el mundo real (In-the-Wild) en algún momento. (Una gran cantidad de códigos maliciosos, entre los cuales se destacan los virus que a veces llamamos virus zoo, nunca llegan a activarse en el mundo real.)

Las muestras de WildCore, la colección en la que se basan las evaluaciones satisfactorias, ya han pasado por un proceso de validación, aunque aún así se espera que los evaluadores con acceso a ellas vuelvan a generar y validar sus propias muestras en lugar de sólo arrojar las muestras a un analizador. Esta colección ofrece un punto de partida para las evaluaciones comparativas, quizá el mejor que tengamos en la actualidad, por más parcial o imperfecto que sea. Si se parte desde una buena base, se reduce el riesgo de obtener falsos positivos (objetos inocentes calificados de manera errónea como maliciosos): de lo contrario, un producto competente puede ser penalizado por tener razón, porque el evaluador asume en forma incorrecta que no fue capaz de detectar el código malicioso. (Es probable que ya hayamos mencionado este tema...) A pesar de todo, sería ingenuo pretender que las evaluaciones basadas en la lista de WildList son admiradas universalmente, incluso dentro de la comunidad de la industria antivirus [17, 42]. Los problemas con la lista actual WildList son muy conocidos (incluyendo para la organización WildList – <http://www.wildlist.org> –, que en el presente se está ocupando de tratarlos):

- Sólo se incluyen en la lista los códigos maliciosos replicativos.
- WildCore sólo representa una pequeña proporción de todos los códigos maliciosos (incluso de los códigos maliciosos replicativos, aunque puede argumentarse que incluye todos los virus y gusanos que se suelen considerar dentro de los más críticos). Es cierto que esto también puede decirse de cualquier paquete pequeño, pero en mayor medida.
- La lista WildList siempre se encuentra un paso más atrás: los rigurosos requerimientos para validar las muestras antes de agregar cada variante generan una demora significativa incluso para la organización con recursos más completos.
- Sin duda alguna, los evaluadores que no se encuentran dentro del círculo privilegiado también señalarán las dificultades para ser aceptados como receptores de WildCore: esto refleja que existe una clara necesidad de confiar en la competencia y la autenticidad del receptor, pero sigue siendo una fuente de discordia.

A pesar de todo, las evaluaciones que usan la lista WildList siguen siendo un componente esencial en algunas pruebas actuales [43], probablemente por las siguientes razones:

- Es razonable que las muestras de WildCore se consideren códigos maliciosos reales (replicativos), no archivos basura.
- Las muestras ya han sido validadas e identificadas (aunque sigue existiendo la necesidad de que el evaluador efectúe otra validación posterior – o al menos otra replicación).
- Los factores mencionados arriba minimizan el riesgo de efectuar identificaciones erróneas y falsos positivos.
- Proveen una colección consistente como punto de partida.

El hecho de que se espere que la mayoría de los vendedores de las principales corrientes tengan acceso a dichas muestras y las detecten correspondientemente, se suele mencionar como prueba de la insuficiencia de la colección como un criterio de evaluación: sin embargo, el hecho de que con frecuencia haya una gran discrepancia entre los productos en el contexto de una evaluación satisfactoria sí sugiere que es posible aprender algo útil de las evaluaciones con la lista WildList.

Incluso en casos donde la lista WildList no es una base apropiada para la evaluación, evaluaciones de troyanos o quizás evaluaciones basadas en un sistema más proactivo, se espera que se sigan los mismos estándares estrictos para la validación de muestras.

## Detectar la falacia

En septiembre del 2007, la revista SC Magazine informó que “El informe tan esperado (...) de la Cámara de los Lores (la Cámara Alta del Parlamento de Reino Unido) (...) [recomienda] (...) incrementarles el pasivo a los vendedores de servicios de seguridad en tecnología de la información que generen fallas en seguridad (...) Tanto McAfee como Symantec señalaron la complejidad de la industria de tecnología de la información y el potencial de los usuarios para comprometer productos que, de lo contrario, serían seguros”. Incluyeron una cita de un vocero de McAfee, que expresó que: “Sería muy difícil hacer responsables a los vendedores por las fallas en seguridad, ya que todo es consecuencia de la manera en que se realiza el despliegue de la solución. Un vendedor de seguridad provee herramientas para las empresas, pero es responsabilidad de las empresas usarlas en forma correcta”.

Ninguna de estas afirmaciones es incorrecta, o al menos estamos seguros de que no tienen la intención de engañar al público (por ejemplo, McAfee especifica en algunas publicidades que su programa “no garantiza la protección contra todas las amenazas posibles”). Pero deja a los consumidores con la idea ya bastante común de que, si no desestabilizan la configuración del programa, estarán totalmente protegidos. Y para la mayoría de las personas, eso significa que todos los códigos maliciosos serán detectados. Como es de esperar, esto no es cierto, y se convierte en una de las razones por las que las personas piensan lo peor de nosotros. Cada falso negativo (por no mencionar cada falso positivo importante) se considera un fracaso reprochable en el suministro de un nivel de protección que los códigos maliciosos conocidos e incluso los analizadores heurísticos no son capaces de lograr en el mundo real con la situación actual de amenazas. Sin duda, los vendedores honestos nunca han prometido una protección del 100% usando exploradores basados principalmente en firmas (o sea, firmas de virus específicos y firmas heurísticas): aquí estamos en presencia de un pensamiento iluso. Lo que los consumidores quieren en realidad, es una identificación automática pseudoexacta de todas las amenazas (a diferencia de las soluciones genéricas), que no requieran ninguna toma de decisiones por parte de ellos.

En esta ocasión no trataremos el tema de que si uno logra hacer ajustes menores en la configuración predeterminada del programa tal cual salió de la caja, por lo general logrará mejorar rotundamente la seguridad.

¿Qué tiene que ver todo esto con la evaluación? Simplemente lo siguiente: si no podemos confiar en el producto antimalware para procesar las amenazas y protegernos del amplio espectro de amenazas que ingresan, y tampoco podemos identificar y verificar todas las amenazas que encontraremos en algún lugar del ciberespacio en el momento de la evaluación, lo mejor que podemos tratar de hacer es tomar una “fotografía” representativa de la escena actual de amenazas con la cual podamos ejecutar los

productos en evaluación. Entonces, es de incumbencia para los evaluadores poner todo su esfuerzo en asegurar que la fotografía en cuestión se asemeje lo más posible a la topología real de ese panorama de amenazas. Una fotografía tomada al azar donde aparezcan una o dos rocas – o los virus que se esconden debajo – no reflejará la topología con la suficiente precisión como para funcionar sobre una base satisfactoria que llegue a conclusiones satisfactorias.

## ¿Cuán práctica es la evaluación “hágalo usted mismo”?

La evaluación de una buena detección requiere, entre otras cosas, procedimientos meticulosos y colecciones de muestras extensas, mantenidas cuidadosamente, tanto de malware como de archivos limpios (para la evaluación de falsos positivos) sin incluir archivos defectuosos. Es un proceso muy intensivo, ya que consume tiempo y recursos, es demandante en un nivel técnico y es difícil de lograr sin la cooperación de la industria donde las muestras se comparten entre individuos de confianza. Los gastos involucrados significan que los resultados no estarán disponibles para quienes no estén subscriptos.

En un paquete validado para la evaluación, las muestras defectuosas son eliminadas minuciosamente y se usan técnicas suplementarias como los paquetes de muestras analizadas con detenimiento para las pruebas de falsos positivos. Esos paquetes de falsos positivos deben ser cuidadosamente clasificados. Piense en un archivo de auto extracción o en un archivo comprimido: no es lo mismo que un archivo ejecutable, ya que contiene objetos múltiples. Algunos programas antimalware analizarán todos los objetos incluidos, otros sólo analizarán el objeto que los contiene; la argumentación radica en que si se ejecuta un archivo malicioso, será atrapado en ese punto. Estos tipos de archivos deben clasificarse con cuidado para lograr una evaluación consistente de detección de falsos positivos exactos (“manzanas por manzanas”). Más tarde, cuando se hace la evaluación real, los procedimientos se planifican, documentan y cumplen con precisión. A menudo, el servicio está financiado por los vendedores cuyos productos se están evaluando, con frecuencia en desventaja (ya sea en forma intencional o no) de los pequeños vendedores y proyectos comunitarios. De todas formas, los resultados completos deben poder ser reproducidos, por lo que se almacenan en forma apropiada, y todos los datos de la evaluación se recogen y archivan en caso de que más tarde se cuestione el método empleado.

Muchas de las recomendaciones comunicadas en los círculos de seguridad son informales, basadas en el supuesto rendimiento sin inconvenientes de una instalación efectuada en vivo. Las variables como la configuración y la calidad de los paquetes de evaluación utilizados deben ser tomadas por confiables, dada la ausencia de una metodología de evaluación con informes claros.

Al menos en este punto podemos felicitar la tentativa de Dirk Morris de Untangled [1], que intentó explicar la metodología que él empleó, aunque no fue capaz de responder a preguntas específicas. La ardua consecuencia (por no decir injusta) fue que, por ser honesto – lo que es elogiado (e ingenuo) – en cuanto a mostrar sus métodos y su paquete de muestras, fue más fácil criticar los agujeros evidentes en su metodología. A pesar de ello, la estrategia de transparencia ha sido útil en varios casos, ya que entidades evaluadoras tan admiradas y respetadas como Virus Bulletin varias veces publicaron reacciones o cambios cuando se encontraron problemas genuinos, lo que en ocasiones ha servido para modificar la metodología en el futuro.

## Los virus no son todo el problema

No es que no sean malos cuando nos llega uno, pero son sólo un porcentaje (cada vez menor) de la amplia variedad de códigos maliciosos. En consecuencia, la elección de una solución antimalware se ve afectada por un abanico completo de elementos secundarios de detección; es decir, cuán efectiva es al detectar códigos maliciosos no virales (en oposición a los objetos legítimos) y algunas categorías cada vez mayores de amenazas no completamente maliciosas (diríamos que se encuentran “en la escala de los grises”) como utilidades de administración remota, además de un amplio rango de otras cuestiones como la usabilidad.

Mientras aquí nos centramos en el rendimiento (y en particular en la detección y hasta cierto punto en la desinfección, aunque no se incluye con tanta frecuencia en las evaluaciones formales – quizá tenga que ver con la cantidad cada vez mayor de recursos que requiere) hemos definido una serie de temas principales para la evaluación [37] que no se tratarán todos aquí (sabemos que el orden puede resultar polémico: siempre existirá el compromiso entre “la mejor práctica en seguridad” y “lo que exige el director ejecutivo”):

- Costo
- Rendimiento
- Facilidad de uso
- Alcance de las funciones
- Capacidad de configuración
- Funciones de soporte

No analizaremos las metodologías individuales de evaluación en este momento, pero algunos de los tipos de evaluaciones son los siguientes [38]:

- Evaluación proactiva (retrospectiva o congelada) de capacidades heurísticas
- Evaluación del momento de actualización del producto (a veces llamado evaluación de respuesta)
- Evaluación de códigos maliciosos activos en el mundo real (In-the-Wild)
- Evaluación de colecciones zoo
- Evaluación de códigos maliciosos no replicativos
- Evaluación en tiempo real
- Evaluación bajo demanda
- Evaluación de falsos positivos

De hecho, la evaluación del rendimiento puede conllevar, de manera potencial, un gran abanico de objetivos de detección (y en algunos casos de desinfección) [37], como los siguientes:

- Códigos maliciosos activos en el mundo real (In-the-Wild)
- Virus zoo
- Nuevas amenazas importantes
- Amenazas desconocidas (rendimiento heurístico)
- Diversidad de amenazas detectadas
- Virus de sistema (hardware/firmware/virus específicos del sistema operativo)
- Virus parásitos
- Virus de macro y de líneas de comandos
- Amenazas de partes múltiples/multipolares
- Códigos maliciosos específicos de correos electrónicos (programas que envían mensajes desde una máquina infectada, programas que envían correos electrónicos masivos, etc.)
- Códigos maliciosos que se encuentran en el servidor web
- Gusanos de red, híbridos de gusanos y virus
- Troyanos (destructivos, ladrones de contraseñas, programas de puerta trasera, troyanos bancarios, etc.)
- Bots
- Virus latentes
- Virus de plataformas múltiples o de transmisión heterogénea
- Generadores
- Virus fallados, corruptos, otros archivos no viables
- Bromas
- Programas espía
- Propaganda no deseada
- Muchas otras amenazas que no son tan conocidas

### ¿Qué se necesita para hacer una evaluación satisfactoria?

- Metodología apropiada y empleada en forma correcta
- Posibilidad de reproducir la evaluación
- Resultados y métodos verificables en forma independiente
- Paquetes de muestras validados y realistas
- Adherencia a prácticas seguras y éticas cuando se manejan y evalúan las muestras
- Comprensión de lo que es (y lo que no es) la tecnología que uno está evaluando

La mayoría de los evaluadores aficionados (muchos los cuales creen ser profesionales en seguridad) no comprenden la necesidad de las metodologías de evaluación satisfactorias (separación de objetivos, configuración consistente) y la necesidad de entender sobre metodologías antimalware (en particular, las tecnologías de detección); de ahí surgen los numerosos informes que confunden la evaluaciones de la identificación exacta o casi exacta, la heurística y las tecnologías genéricas como el uso de listas blancas.

Gran parte de la evaluación se basa en el intento de “engañar” a los exploradores [38]; por ejemplo al ejecutarlos con muestras contextualizadas o modificadas de manera inapropiada. Nosotros creemos que esta estrategia es sospechosa desde el punto de vista ético, en especial por la forma en que puede confundir a la audiencia. La evaluación de falsos positivos, por ejemplo, requiere un paquete apropiado de muestras de falsos positivos activos en el mundo real (es decir, objetos de prueba que realmente se puedan encontrar en computadoras, no muestras falsas diseñadas especialmente para la ocasión). Los archivos “grises”, inusuales o muy raros y poco probables de ser encontrados siempre tienden a penalizar los productos basados en heurística que marcan objetos que no parecen “normales”.

Los paquetes insuficientes de muestras que contienen archivos basura sólo sirven para confundir más las cosas: cuanta más basura se agregue a los paquetes de evaluación, los analizadores se verán en la necesidad de detectar más objetos irrelevantes simplemente para poder seguir en juego.

La evaluación del “momento de actualización del producto” suele introducir una preferencia estadística, donde los recursos de los productos más exitosos se evalúan con menos muestras y es menos apropiado para la evaluación comparativa que para la evaluación proactiva. El foco en la velocidad de la actualización envía un mensaje erróneo a los consumidores, dándoles la falsa impresión de que un producto que envía una gran cantidad de actualizaciones muy seguidas, tiene mejor nivel de protección.

La evaluación retrospectiva (proactiva), donde las actualizaciones se congelan por un período preestablecido de tiempo, si está bien administrada, es una evaluación heurística mucho mejor que las estrategias que usan kits de virus, muestras hechas a medida, etc. Igualmente no es una técnica sencilla.

## ¿Quiénes son los evaluadores confiables?

Las evaluaciones realizadas por ciertas organizaciones que siguen lineamientos bastante uniformes, en general son consideradas válidas por la comunidad antivirus (notoriamente conservadora) y surgen de la necesidad de implementar un conjunto de metodologías imparciales que sirvan como punto de partida. Para ello se requiere disponer de tiempo y de una habilidad considerable, y ese gasto es una de las razones por las que los resultados de muchas evaluaciones de primera categoría, en especial su metodología completa, no están disponibles en forma gratuita, es decir, sólo están disponibles para los subscriptores, al menos en el corto plazo.

Por más irritados que estén los vendedores e investigadores por la necesidad y en especial, la implementación de (la mayoría de) las evaluaciones comparativas, en general terminan admitiendo con reticencia la necesidad que tienen los clientes de contar con alguna información comparativa. Ninguno de los siguientes grupos tiene la aprobación universal e incuestionable de la comunidad antivirus completa, pero son tomados seriamente:

- Virus Bulletin (<http://www.virusbtn.com>)
- ICSA Labs (<http://www.icsalabs.com>)
- West Coast Labs (<http://westcoastlabs.org>)
- AV-Test.org (<http://www.av-test.org>)
- AV Comparatives (<http://www.av-comparatives.org>)

En forma comparativa, los informes de revistas de informática y otros recursos no especializados son un modo fortuito de evaluar la efectividad de productos antimalware. Son pocos los periodistas no especializados que poseen conocimientos técnicos sobre el tema, en lo que respecta a comprender tanto las tecnologías de las amenazas como las medidas para contrarrestarlas, o sobre las trampas presentes en la evaluación de la detección. Rara vez se describe la metodología de evaluación, en especial si la evaluación es realizada por un tercero. Contratar a otra empresa para hacer la evaluación es una forma muy responsable de lidiar con ella, siempre y cuando la organización evaluadora sea competente [2].

Los informes pueden centrarse en aspectos más subjetivos como la usabilidad, el impacto en los recursos del sistema, la velocidad percibida, entre otros. Este enfoque puede resultar problemático [45], ya que los temas que les conciernen a los administradores de sistemas o gerentes y directores de seguridad quizá no sean evidentes para una persona ajena a dichos cargos o para cualquiera que piense en función de computadoras individuales en el hogar o la oficina pequeña. De todas formas, estos aspectos:

- Son más susceptibles a las evaluaciones inexpertas
- Tienen menos tendencia a generar consecuencias serias cuando se realizan de manera incompetente

Lamentablemente, siguen existiendo evaluaciones donde se sospecha que la elección del editor está injustamente influenciada por el listado de anunciantes. En un artículo del doctor Alan Solomon [24], se describen diversas formas por las que los resultados de evaluaciones comparativas han reflejado de manera casual o deliberada la parcialidad o la agenda comercial del evaluador. Por desgracia, a pesar de la antigüedad del artículo, los principios generales y algunos detalles específicos siguen siendo tan relevantes en la actualidad como lo eran en la década de los '90.

Las revistas especializadas como Virus Bulletin y las organizaciones evaluadoras de buena reputación, como las instalaciones para realizar evaluaciones en Magdeburg, tienden a ofrecer información más confiable; pero estas organizaciones suelen enfocarse más que nada en la detección (incluyendo variaciones como la evaluación de falsos positivos), dejando de lado la gama completa de características. Las cuestiones como la usabilidad son muy importantes y es un área que los evaluadores profesionales rara vez tratan en detalle; en especial porque en realidad, desde el punto de vista conceptual y práctico, la detección es más fácil de evaluar que la usabilidad si uno cuenta con los recursos y los conocimientos para hacerlo apropiadamente.

## Configuración predeterminada

Nos gustaría tratar un último tema importante que se suele usar para justificar metodologías donde no se hace ningún intento para nivelar el campo de juego: en efecto, todas las evaluaciones de detección de esas pruebas se basan en configuraciones predeterminadas, tal cual estaba el producto al salir de su caja.

Un usuario final no va a usar necesariamente una configuración que atrape todas las muestras y la mayoría de las configuraciones predeterminadas priorizan la velocidad por sobre el análisis. Entonces existe una gran distinción entre la capacidad de detección predeterminada y la de detección completa (aquí también hay un problema con los niveles de configuración de la heurística). Si un producto es capaz de detectar 100.000 cepas de virus, pero no en la configuración predeterminada, y además el vendedor le dificulta al consumidor que lo use con la configuración que le proporcionaría mayor provecho, allí surge un problema para la evaluación de la usabilidad y la configuración: no es una evaluación de detección total. No obstante, ciertamente existe un argumento para evaluar la detección predeterminada (aunque en ese caso uno quizá también deba evaluar con la configuración de máxima seguridad). Sin embargo, es más difícil evaluar los programas en la configuración predeterminada porque la cantidad de variables dificulta el establecimiento de la paridad entre las configuraciones evaluadas: por el contrario, uno no sólo evalúa el rendimiento sino también la filosofía de la configuración. Pero eso no significa que no vale la pena intentarlo.

No obstante, existen muchas evaluaciones intermedias que conviene tener en cuenta como aquellas para medir distintos niveles de heurística, de análisis bajo demanda comparados con análisis en el acceso, etc. además de evaluaciones estrictamente limitadas como la susceptibilidad ante los archivos para evaluaciones, (en particular el archivo EICAR), la desinfección de virus de macro, etc.

## Conclusión

Quizás uno no necesite ser un investigador de antivirus para poder evaluar un programa antivirus, a pesar de que la evaluación es un campo muy específico dentro del de la investigación antivirus. Algunas de las reglas para evaluar programas antivirus en el nivel del consumidor son las mismas que para otros tipos de productos, pero es más complicado porque todos tienen una idea general de cómo usar, por dar un ejemplo, un procesador de texto y qué se puede esperar de él. En cambio, muchas personas tienen una idea bastante distorsionada de lo que hace un antivirus y cómo lo hace. Nosotros creemos que la comunidad de investigación de programas antivirus fue la causante de este estado desafortunado de las cosas, pero ya entraríamos en otro debate.

No pretendemos que las personas tomen todo lo que nosotros, o la industria antimalware en general, decimos como si fueran leyes escritas en tablas de piedra. Estamos todos a favor del sano escepticismo. Lo que nos resulta inadmisibles es la tendencia a asumir que la industria antivirus es un gran fraude y que cualquier idea que no proceda de ese sector es, por ende, verdadera.

No pretendemos que las personas tomen todo lo que nosotros, o la industria antimalware en general, decimos como si fueran leyes escritas en tablas de piedra. Estamos todos a favor del sano escepticismo. Lo que nos resulta inadmisibles es la tendencia a asumir que la industria antivirus es un gran fraude y que cualquier idea que no proceda de ese sector es, por ende, verdadera.

Son muchos los problemas para facilitarles la evaluación a personas ajenas a la industria y no tenemos las soluciones para todos ellos. Enviar muestras a personas en las que uno no sabe si puede confiar es un área problemática comprensible. Existen soluciones parciales para este problema: contratar a una empresa competente para realizar la parte de detección de la evaluación comparativa o usar los recursos que una entidad de ese tipo (o un vendedor de antimalware) ponen a disposición del público bajo condiciones estrictamente controladas para que las muestras no se “escapen”, por ejemplo. Sin embargo, lograr que las personas sean más conscientes de las prácticas buenas y malas, enseñarles lo que pueden y lo que no pueden hacer, darles el poder para realizar sus propias evaluaciones significativas y juzgar las evaluaciones ajenas es (esperamos y creemos) un paso práctico hacia un entendimiento y una práctica superiores.

Creemos que la industria antivirus tiene la obligación de tratar este tema mejor de lo que lo ha hecho hasta ahora y, desde nuestro pequeño lugar, esperamos tratar el problema en un libro entero en un futuro próximo.

## Referencias

- [1] <http://blog.untangle.com/?p=95>; <http://blog.untangle.com/?p=96>
- [2] David Harley: "AV Testing SANS Virus Creation," Virus Bulletin, October 2006.
- [3] Igor Muttik: "Shall we all write viruses to find the best antivirus?" en <http://www.avertlabs.com/research/blog/?p=71>
- [4] Igor Muttik: "A Tangled Web", en "AVIEN Malware Defense Guide for the Enterprise," Syngress 2007
- [5] Alex Eckleberry: "More Testing Silliness" en <http://sunbeltblog.blogspot.com/2006/08/more-testing-silliness.html>
- [6] David Harley: "Untangling the Wheat from the Chaff in Comparative Anti-Virus Reviews" en [http://www.smallblue-greenworld.co.uk/AV\\_comparative\\_guide.pdf](http://www.smallblue-greenworld.co.uk/AV_comparative_guide.pdf)
- [7] David Harley: "Insider's Guide to Comparative Anti-Virus Reviews" en [http://blogs.technet.com/industry\\_insiders/pages/insider-s-guide-to-comparative-anti-virus-reviews.aspx](http://blogs.technet.com/industry_insiders/pages/insider-s-guide-to-comparative-anti-virus-reviews.aspx)
- [8] Vesselin Bontchev: "About Anti-Virus Testing" en <http://www.fprot.com/workshop2007/presentations.html>
- [9] David Harley: "I'm OK, You're Not OK" en "Virus Bulletin", November 2006 (ver <http://www.virusbtn.com/virusbulletin/archive/2006/11/vb200611-OK.dkb>)
- [10] David Harley et al: "Customer Power & AV Wannabes" en "AVIEN Malware Defense Guide for the Enterprise", Syngress 2007.
- [11] David Harley: "Fact, Fiction and Managed Anti-Malware Services" en "Proceedings of the 13th Virus Bulletin International Conference" (2003)
- [12] Justin Kruger y David Dunning "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments" en "Journal of Personality and Social Psychology" Volumen 77 No. 6 (1999) páginas 121-1134
- [13] David Harley, Jimmy Kuo: "Can I get a virus to test my antivirus with?" en alt.comp.virus FAQ, <http://www.faqs.org/faqs/computer-virus/alt-faq/part4/>
- [14] Martin Overton: "FAT32 – New Problems for Anti-Virus, or Viruses" en "Proceedings of Virus Bulletin Conference, October 1997."
- [15] Andrew Hayter: "Nature of Anti-Malware Testing and Certification Programs Life and times of testing Anti-virus Products" para AVAR 2007.
- [16] Maik Morgenstern y Andreas Marx, AV-Test.org: "Testing of 'Dynamic Detection'" para AVAR 2007

- [17] Vesselin Bontchev: "Maintaining a Malware Collection" en <http://www.fprot.com/workshop2007/presentations.html>
- [18] Vesselin Bontchev: "Analysis and Maintenance of a Clean Virus Library" en <http://www.people.frisk-software.com/~bontchev/papers/virlib.html>
- [19] J.B. Rhine, "New Frontiers of the Mind", Farrar and Rhinehart 1937
- [20] <http://www.informatik.uni-hamburg.de/AGN/vtc/>
- [21] <http://www.av-test.org/>
- [22] Henk K. Diemer, David Harley: "Perilous Outsorcery" en "AVIEN Malware Defense Guide for the Enterprise", Syngress 2007
- [23] Tim Wilson: "Antivirus Tools Underperform When Tested in LinuxWorld 'Fight Club'" [http://www.darkreading.com/document.asp?doc\\_id=131246&WT.svl=news1\\_5](http://www.darkreading.com/document.asp?doc_id=131246&WT.svl=news1_5)
- [24] Dr. Alan Solomon: "A Reader's Guide to Reviews" (originalmente publicado en "Virus News International" y atribuido a Sarah Tanner), en [www.softpanorama.org/Malware/Reprints/virus\\_reviews.html](http://www.softpanorama.org/Malware/Reprints/virus_reviews.html)
- [25] Pamela Kane: "The Dangers of Experts" en "PC Security and Virus Protection Handbook", M&T Books, 1994
- [26] Andreas Marx and Frank Dessmann: "The WildList is Dead, Long Live the WildList" en "Proceedings of the 17th Virus Bulletin Conference" 2007
- [27] Randy Abrams: "AV Industry Comments on Anti-Malware Testing" en Virus Bulletin, June 2007
- [28] Mary Landesman: "The Wild WildList" en Virus Bulletin, July 2007
- [29] Rob Rosenberger: "False Authority Syndrome", en <http://www.cknow.com/vtutor/FalseAuthoritySyndrome.html>
- [30] <http://www.sans.org/newsletters/newsbites/newsbites.php?vol=8&issue=65>
- [31] Sarah Gordon y Richard Ford: "When Worlds Collide: Information Sharing for the Security and Anti-Virus Communities", en "Proceedings of the Virus Bulletin Conference" 1999.
- [32] Respuestas a un artículo escrito por John Leyden [http://www.theregister.co.uk/2007/09/28/nsa\\_hacker\\_malware\\_defense\\_project/comments/#c\\_68630](http://www.theregister.co.uk/2007/09/28/nsa_hacker_malware_defense_project/comments/#c_68630)
- [33] Bruce Schneier: "Teaching Viruses" en <http://www.schneier.com/crypto-gram-0706.html#5>
- [34] John Aycok y Alana Maurushat: "Future Threats" en "Proceedings of the 17th Virus Bulletin International Conference" 2007.
- [35] Hiep Dang: "What a Tangled Web" en <http://www.avertlabs.com/research/blog/index.php/2007/08/12/what-a-tangled-web/>
- [36] Tim Wilson: "Antivirus Tools Underperform When Tested in LinuxWorld 'Fight Club'" [http://www.darkreading.com/document.asp?doc\\_id=131246&WT.svl=news1\\_5](http://www.darkreading.com/document.asp?doc_id=131246&WT.svl=news1_5)

- [37] David Harley y Robert Slade: "Product Evaluation and Testing" en "Viruses Revealed" (Harley, Slade, Gattiker), Osborne 2001.
- [38] David Harley y Andrew Lee: "Antimalware Evaluation and Testing", en "AVIEN Malware Defense Guide for the Enterprise" Syngress 2007;
- [39] Dirk Morris: "Selling Dead Donkeys" en <http://blog.untangle.com/?p=20>
- [40] Joe Wells: "Pragmatic Anti-Virus Testing" en Virus Bulletin, September 2001. También disponible en <http://www.sunbeltsoftware.com/ihs/alex/Pragmaticantivirustesting.pdf>
- [41] Sarah Gordon: "What is Wild?" en <http://csrc.nist.gov/nissc/1997/proceedings/177.pdf>
- [42] Vesselin Bontchev: "About Anti-Virus Testing" en <http://www.f-prot.com/workshop2007/presentations.html>
- [43] <http://www.virusbtn.com/vb100/about/100procedure.xml>;  
[http://www.icsalabs.com/icsa/topic.php?tid=453f\\$2571e0c1-26134a78\\$461e-02308865](http://www.icsalabs.com/icsa/topic.php?tid=453f$2571e0c1-26134a78$461e-02308865)
- [44] Andrew Lee: "Testing Heuristics" en <http://www.f-prot.com/workshop2007/presentations.html>
- [45] Igor Muttik: "Comparing the Comparatives" en [http://www.mcafee.com/us/local\\_content/white\\_papers/threat\\_center/wp\\_imuttik\\_vb\\_conf\\_2001.pdf](http://www.mcafee.com/us/local_content/white_papers/threat_center/wp_imuttik_vb_conf_2001.pdf)

## Recursos adicionales

- Sarah Gordon y Richard Ford: "Real World Anti-Virus Product Reviews And Evaluations – The Current State Of Affairs" en <http://csrc.nist.gov/nissc/1996/papers/NISSC96/paper019/final.PDF>
- Adam J. O'Donnell: "Real-World Testing of Email Anti-Virus Solutions", en Virus Bulletin, marzo de 2007
- [http://www.av-comparatives.org/seiten/ergebnisse\\_2007\\_02.php](http://www.av-comparatives.org/seiten/ergebnisse_2007_02.php)
- <http://www.av-comparatives.org/seiten/ergebnisse/2ndgrouptest.pdf>
- Randy Abrams: "AV Industry Comments on Anti-Malware Testing" en Virus Bulletin, junio de 2007.
- Igor Muttik: "Antivirus Testing Workshop in Reykjavik" en <http://www.avertlabs.com/research/blog/index.php/2007/05/29/antivirus-testing-workshop-in-reykjavik/>
- Richard Ford, Attila Ondi: "Testing Times Ahead?", Virus Bulletin, abril de 2007
- Randy Abrams: "Doesn't the EICAR test file look spiffy?" en <http://www.eset.com/threat-center/blog/?p=15>
- Randy Abrams: "Giving the EICAR Test File Some Teeth" en el evento "Ninth International Virus Bulletin Conference and Exhibition", 1999