ESET White Paper

¿ES POSIBLE QUE LA INTELIGENCIA ARTIFICIAL POTENCIE EL MALWARE EN EL FUTURO?

Ondrej Kubovič – Especialista en concientización sobre seguridad de ESET

Con la contribución de

Peter Košinár – Asistente técnico de ESET
Juraj Jánošík – Ingeniero de software senior de ESET

ENJOY SAFER TECHNOLOGY™

CONTENIDO

INTRODUCCIÓN

La inteligencia artificial (IA) es un tema muy recurrente estos días. Es el eslabón principal en los argumentos de venta, "potencia" varios servicios online y se menciona en casi todos los productos nuevos al buscar inversores. Mientras que algunos proveedores realmente tratan de incorporar en sus productos el valor de esta tecnología para beneficio de sus clientes, otros solo la usan como una palabra de moda, pero no son capaces de cumplir sus promesas.

Una simple búsqueda online de sus siglas en inglés ("AI") hoy arroja casi 2.200 millones de resultados, lo que demuestra el gran interés de tanto los expertos como el público en general. Parte de la moda se puede atribuir a las grandes hazañas logradas gracias a esta tecnología (permitirles a los investigadores ver a través de las paredes), aunque también tiene connotaciones negativas, por ejemplo, se predice que la IA podría eliminar millones de puestos de trabajo y dejar obsoletas industrias enteras.

El aprendizaje automático o Machine Learning (ML), como subcategoría del verdadero objetivo autosustentable de la IA, que aún dista mucho de ser alcanzado, ya ha desencadenado cambios radicales en muchos sectores, incluyendo el de la seguridad cibernética. Las mejoras en los motores de exploración, en la velocidad de detección y en la capacidad de identificar irregularidades fueron factores que contribuyeron a proteger mejor a las empresas, en especial frente a las amenazas nuevas y emergentes, así como las amenazas persistentes avanzadas (APT, del inglés).

Desafortunadamente, esta tecnología no está disponible en forma exclusiva para los defensores de la seguridad. Los hackers de sombrero negro, los ciberdelincuentes y otros actores maliciosos también son conscientes de los beneficios de la IA y probablemente intentarán aplicarla a sus actividades de alguna forma. Es probable que los ataques dirigidos a empresas específicas y el robo de dinero o de datos comiencen a ser más difíciles de descubrir, rastrear y mitigar.

Incluso podríamos argumentar que estamos por comenzar una era en la que los "ataques cibernéticos potenciados por la IA" se convertirán en la norma y reemplazarán a aquellos operados por actores maliciosos altamente calificados. ESET, como proveedor de seguridad establecido que ha estado luchando contra los ciberdelincuentes por décadas, entiende los próximos desafíos y los posibles escenarios futuros, y los explica en el presente documento.

Para proporcionar una visión más amplia, también se incluyen los resultados de una encuesta que ESET le encargó a OnePoll. Las opiniones y preocupaciones sobre el uso de la IA y el ML en el contexto de la seguridad cibernética se midieron en esta encuesta realizada a casi 1000 responsables de la toma de decisiones sobre TI de empresas con más de 50 empleados en los Estados Unidos, el Reino Unido y Alemania.

Para evitar posibles confusiones, en este white paper también se abordan las diferencias entre la IA y el ML, y se detallan los límites de este último.

Finalmente, se ofrece una descripción general de los "ataques basados en la IA", además de información sobre el diseño del motor de Machine Learning de ESET, Augur, y un resumen de sus productos de nivel corporativo especialmente creados aprovechando esta tecnología para combatir constantemente las amenazas cibernéticas emergentes y cambiantes.

Inteligencia artificial vs Machine Learning

La idea de la inteligencia artificial existe desde hace más de 60 años. Representa el ideal aún inalcanzable de una máquina inteligente y autosustentable que puede aprender en forma independiente, basándose solo en las entradas del entorno (por supuesto, sin intervención humana).

Sin embargo, hoy en día se suele usar el término "IA" para referirse solamente a una subcategoría de esta tecnología: a Machine Learning. Este campo de la informática se originó en la década de 1990 y sus aplicaciones en el mundo real les permiten a las computadoras encontrar patrones en grandes

cantidades de datos, analizarlos y actuar sobre los resultados. Estos algoritmos son el ingrediente no tan secreto de todos los productos de seguridad cibernética que mencionan la IA en sus propagandas de marketing.

Supervisado, no supervisado o parcialmente supervisado

En el contexto de la seguridad cibernética, los algoritmos de Machine Learning se utilizan principalmente para clasificar y analizar muestras, identificar similitudes y generar un valor de probabilidad para el objeto procesado, de modo de clasificarlo en una de las tres categorías principales: malicioso, potencialmente no seguro o no deseado (PUSA/PUA, por sus siglas en inglés), o no infectado.

Sin embargo, para lograr los mejores resultados posibles, es necesario "entrenar" esta tecnología usando un conjunto muy extenso de muestras no infectadas y maliciosas correctamente etiquetadas, lo que le permite al algoritmo comprender la diferencia. Esta capacitación y supervisión humana es la razón por la cual se denomina aprendizaje automático supervisado. Durante el proceso de aprendizaje, se le enseña al algoritmo cómo analizar e identificar la mayoría de las amenazas potenciales en el entorno protegido y también cómo actuar de manera proactiva para mitigarlas. La integración de este algoritmo en una solución de seguridad la hace significativamente más rápida y aumenta su capacidad de procesamiento, en comparación con otras soluciones que solo usan el conocimiento humano para proteger los sistemas del cliente.

Los algoritmos que no se entrenan con datos completa y correctamente etiquetados pertenecen a la categoría de aprendizaje automático no supervisado. Son adecuados para encontrar similitudes y anomalías en el conjunto de datos que podrían escapar al ojo humano, pero no necesariamente aprenden cómo separar lo bueno de lo malo (o para ser más precisos, lo no infectado de lo malicioso). En la seguridad cibernética, esta puede ser una característica muy útil para trabajar con grandes conjuntos de muestras etiquetadas. El aprendizaje no supervisado se puede utilizar para organizar los datos en clústers y ayudar a crear grupos más pequeños, pero mucho más consistentes, para entrenar otros algoritmos.

El aprendizaje automático parcialmente supervisado está entre las categorías de aprendizaje supervisado y no supervisado. En el proceso de aprendizaje del algoritmo solo se usan los datos parcialmente etiquetados, y los resultados son supervisados y ajustados por expertos humanos hasta lograr el nivel deseado de precisión. Este enfoque es necesario porque la creación de un conjunto de entrenamiento completamente etiquetado suele ser una tarea laboriosa, costosa y que lleva mucho tiempo. En otros casos, directamente no existen datos etiquetados en forma completa y correcta, por lo que el aprendizaje parcialmente supervisado es la única opción para generar un algoritmo útil. El motor de Machine Learning de ESET, llamado Augur, funciona de manera similar. Se utiliza para clasificar elementos que no formaban parte de su conjunto de entrenamiento y que no estaban previamente etiquetados.

¿LA "IA" ES SOLO UNA MODA MÁS?

Además de su uso científico original, el término *inteligencia artificia*l (1) también es una palabra de moda. Sin embargo, ¿qué tan exagerada es la publicidad? Gracias a los avances significativos en el campo de Machine Learning y su aplicación más amplia en problemas del mundo real, el interés en la IA creció mucho en los últimos años, y alcanzó picos en 2017 y 2018 que no se veían desde la última década.

Esto queda demostrado al analizar las tendencias de búsqueda de los términos "Machine Learning" e "inteligencia artificial" en inglés.

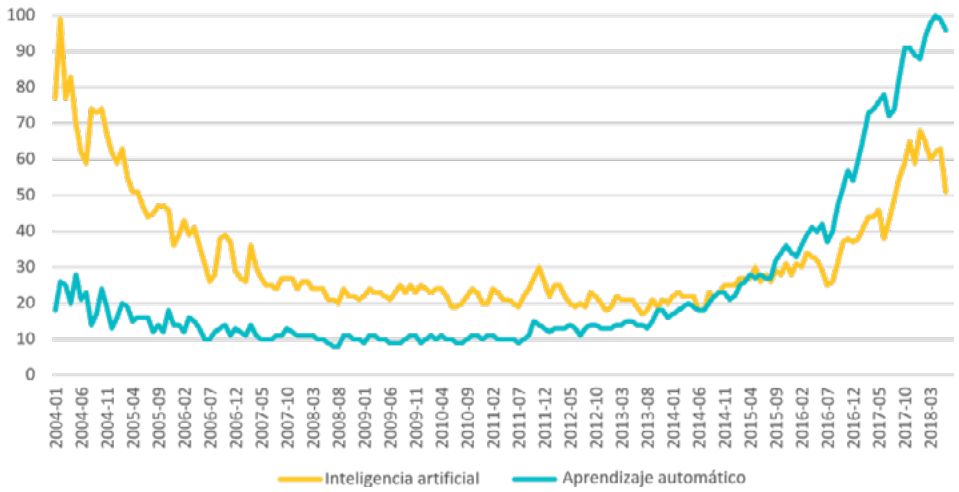

Imagen 1: // Tendencia de búsqueda de los términos "Inteligencia Artificial" y "Aprendizaje Automático" en inglés de 2004 a 2018. Fuente: Tendencias de Google. Esto también aparece en entornos corporativos, donde el aprendizaje automático o la IA parecen estar ampliamente implementados, como se observó en la encuesta que ESET le encargó a OnePoll.

Según los resultados, el 82% de los responsables de la toma de decisiones de TI en empresas estadounidenses, británicas y alemanas con más de 50 empleados creen que su organización ya ha implementado un producto de seguridad cibernética que utiliza Machine Learning. Del resto, el 53% declaró que su organización planea implementar este tipo de solución en los próximos 3-5 años, y el 23% indica lo contrario.

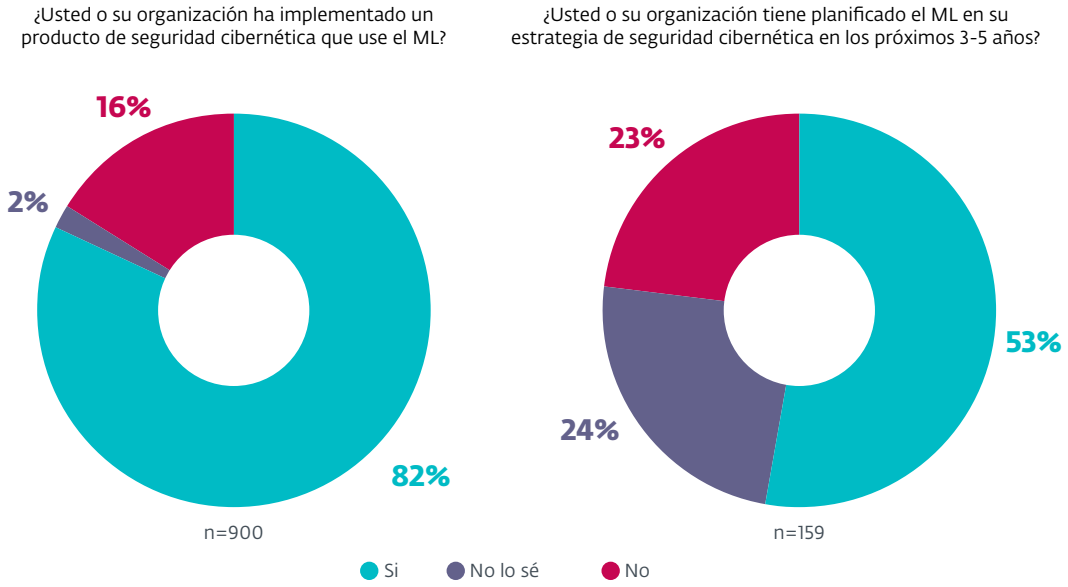

Imagen 2 Imagen 3

El 80% de los encuestados también consideran que la IA y el ML ayudan o ayudarán a su organización a detectar y responder a las amenazas más rápidamente. Los tomadores de decisiones también creen que estas tecnologías los ayudarán a resolver la escasez de personal capacitado en seguridad cibernética para cubrir los puestos de su lugar de trabajo.

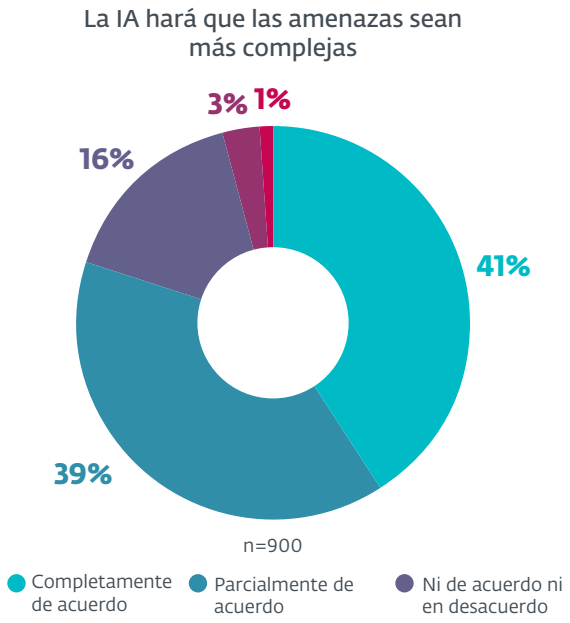

Imagen 4

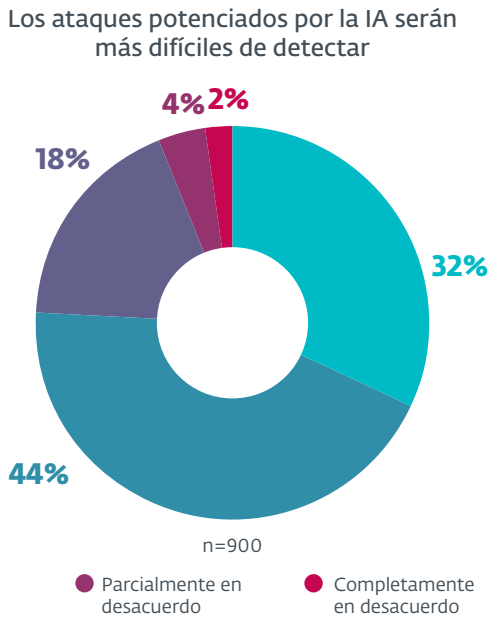

Imagen 5

Con la gran cantidad de marketing en torno a la IA y al ML, muchos de los encuestados llegaron a pensar que estas tecnologías podrían ser la clave para resolver sus problemas más desafiantes de seguridad cibernética; sin embargo, la mayoría también expresó que las afirmaciones sobre la implementación de la IA y el ML en la infraestructura de seguridad les parecían exageradas.

Por lo tanto, sin menospreciar el verdadero valor de la IA y el ML como herramientas útiles en la lucha contra el delito cibernético, hay ciertas limitaciones que deben tenerse en cuenta, por ejemplo, que confiar en una sola tecnología es un riesgo que puede tener consecuencias perjudiciales, en particular cuando el atacante tiene la motivación, el respaldo financiero y el tiempo suficiente para lograr evadir un algoritmo de ML protector. Un enfoque más seguro y equilibrado para la seguridad cibernética corporativa es desplegar una solución de varias capas capaz de aprovechar el poder y el potencial de la IA y el ML, pero con el respaldo de otras tecnologías de detección y prevención.

¿LA IA PUEDE POTENCIAR LOS ATAQUES CIBERNÉTICOS FUTUROS?

Los avances tecnológicos del aprendizaje automático les ofrecen un enorme potencial de transformación a los defensores de la seguridad cibernética. Desafortunadamente, no son los únicos: los ciberdelincuentes también son conscientes de los nuevos beneficios. La encuesta de OnePoll a casi 1000 participantes, entre gerentes y personal de TI responsables de la seguridad de empresas estadounidenses, británicas y alemanas, arrojó que:

Dos tercios (66%) estuvieron de acuerdo en forma parcial o total en que las nuevas aplicaciones basadas en la IA incrementarán el número de ataques a su organización. Una cantidad incluso mayor de encuestados consideran que las tecnologías basadas en la IA harán que las amenazas sean más complejas y más difíciles de detectar (69% y 70% respectivamente).

La IA incrementaría la cantidad de ataques que mi organización tendrá que detectar y resolver

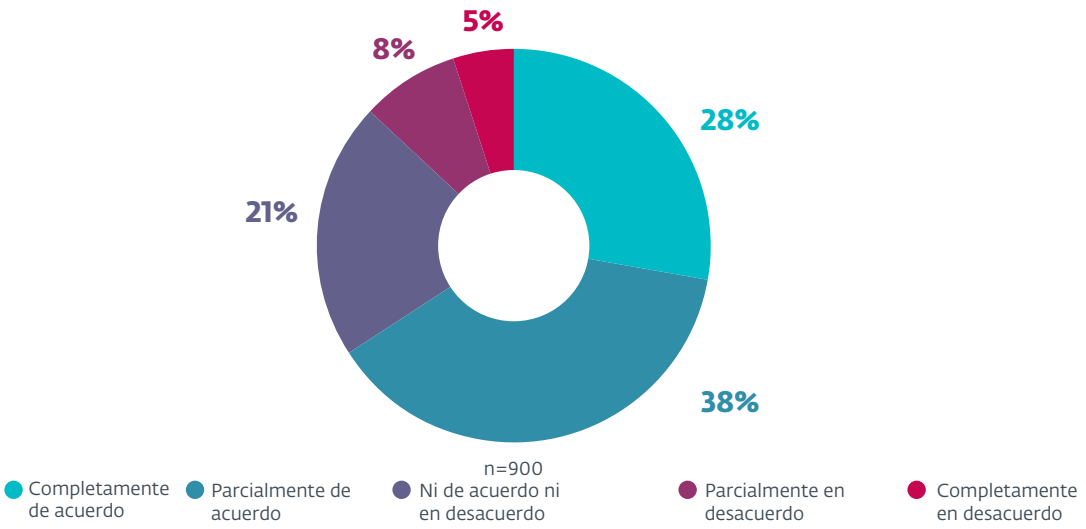

Imagen 6

La IA hará que las amenazas sean más complejas

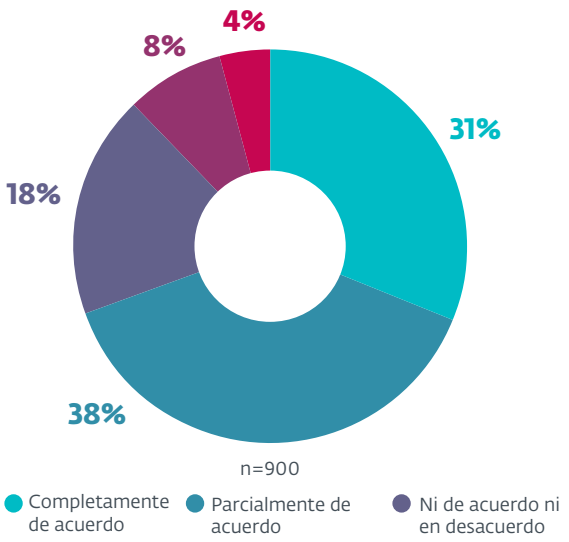

Imagen 7

Los ataques potenciados por la IA serán más difíciles de detectar

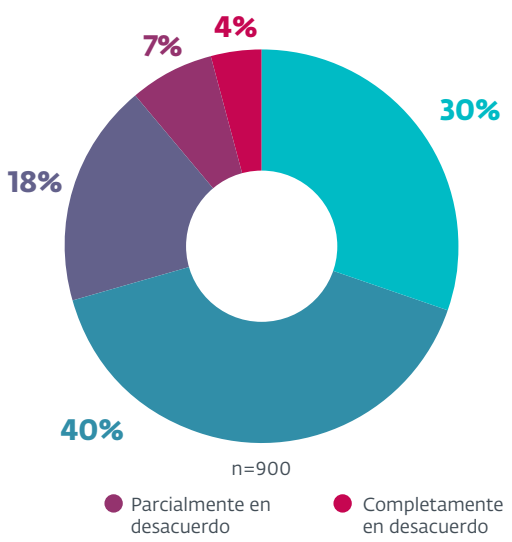

Imagen 8

Aún no sabemos si estas preocupaciones se van a materializar ni cómo. Sin embargo, no sería la primera vez que los atacantes utilizan los avances tecnológicos para ampliar el alcance de sus esfuerzos maliciosos. Ya en 2003, el *troyano Swizzor* (2) utilizó la automatización para volver a empaquetar el malware una vez por minuto. Como resultado, cada víctima recibía una variante polimórficamente modificada del malware, lo que complicaba su detección y permitía su mayor propagación.

De todas formas, este enfoque no sería tan efectivo para evadir las soluciones antimalware modernas, como los productos de ESET para endpoints, que detectan el "ADN" del malware y además pueden identificarlo mediante sus detecciones de red. Sin embargo, mediante el uso de algoritmos avanzados de Machine Learning, los atacantes podrían usar un mecanismo similar al de Swizzor e intentar mejorar su estrategia. Si esta tecnología no es parte de la medida defensiva, el algoritmo del atacante podría

aprender cuáles son los límites de la solución protectora y alterar el código malicioso lo suficiente como para atravesar sus defensas.

Las variantes automáticas del malware distan mucho de ser la única aplicación maliciosa posible de los algoritmos de Machine Learning. A continuación, analizamos algunas de las áreas en las que el uso de esta tecnología podría darles una ventaja a los atacantes (y añadimos algunos ejemplos ilustrativos).

La IA como herramienta para los atacantes

Los atacantes podrían utilizar la IA o el ML para:

Proteger su infraestructura, por ejemplo:

- Detectando intrusos (es decir, investigadores, defensores, cazadores de amenazas) en sus sistemas
- Detectando nodos inactivos y por lo tanto sospechosos en su red

Generar y distribuir nuevos contenidos, tales como:

- Correos electrónicos de phishing
  Diseñados y modificados en forma total o parcial por el algoritmo
- Spam de alta calidad
  Gracias al Machine Learning, la creación de spam nuevo de alta calidad también sería posible para los idiomas menos frecuentes, según la cantidad de material de entrenamiento disponible
- Desinformación
  Combinando automáticamente información legítima con desinformación, y aprendiendo qué es lo que funciona mejor y lo que más comparten las víctimas

Identificar patrones recurrentes, rarezas o errores en el contenido generado y ayudar a los atacantes a eliminarlos

Identificar posibles señales de alarma que probablemente buscarán los defensores

Crear señales falsas para desviar la atención a otros actores o grupos

Elegir el mejor objetivo para su ataque o repartir varias tareas entre distintas máquinas infectadas de acuerdo con su rol dentro de la red, sin la necesidad de una comunicación saliente

Utilizar el modelo de la IA de la **solución de seguridad como una caja negra con intenciones maliciosas**

Los atacantes pueden instalar la solución de seguridad de la víctima en su propio dispositivo con la misma configuración y usarla para identificar qué tipo de tráfico o contenido atravesará sus defensas

Encontrar la técnica de ataque más efectiva

Las técnicas de ataque pueden abstraerse y combinarse para identificar los enfoques más efectivos. Luego, estos enfoques se pueden priorizar para aprovecharlos en el futuro. En caso de que los defensores detecten y anulen a uno de los vectores de ataque, el atacante solo necesitará reiniciar el algoritmo y, basándose en esta nueva información, la tecnología seguirá una ruta de aprendizaje diferente.

Encontrar nuevas vulnerabilidades

Combinando el enfoque anterior con el fuzzing (es decir, proporcionarle al algoritmo datos no válidos, inesperados o aleatorios como entradas), la IA podría aprender una rutina para encontrar nuevas vulnerabilidades.

La IA en el malware

Los desarrolladores de malware podrían utilizar la IA para:

Generar nuevas variantes de malware difíciles de detectar

Como ya se describió en este documento, algunas familias de malware antiguas (por ejemplo, Swizzor) utilizaban la automatización para generar nuevas variantes de sí mismas cada minuto. Esta técnica podría reinventarse y mejorarse usando algoritmos de Machine Learning para aprender cuáles de las variantes recién creadas son las menos susceptibles de ser detectadas y así producir nuevas cepas con características similares.

Ocultar su malware en la red de la víctima

El malware puede monitorear el comportamiento de los nodos o las endpoints en la red objetivo y generar patrones que se asemejen al tráfico de red legítimo.

Combinar varias técnicas de ataque para encontrar las opciones más efectivas que no puedan detectarse fácilmente y priorizarlas sobre alternativas menos exitosas

Ajustar las funcionalidades o el enfoque del malware en función del entorno
Si por ejemplo los ciberdelincuentes quieren atacar navegadores, en lugar de incluir una lista completa de navegadores y posibles escenarios en el malware, solo necesitarán implementar algunos de ellos para las marcas más frecuentes. El algoritmo de IA utilizará esta capacitación y aprenderá directamente en la endpoint a infiltrarse también en los navegadores menos populares y no especificados previamente.

Implementar un mecanismo autodestructivo en el malware que se active si detecta un comportamiento extraño

Al detectar un inicio de sesión de un perfil de usuario o de un programa que no sean los estándar, el malware activa automáticamente el mecanismo de autodestrucción para evitar ser detectado o analizado.

Detectar un entorno sospechoso

Si el algoritmo detecta una máquina virtual, un modo sandbox o alguna otra herramienta utilizada por los investigadores de malware, puede alterar el comportamiento o detener temporalmente su actividad para evitar la detección.

Aumentar la velocidad del ataque

La velocidad de un ataque puede ser crucial, especialmente en casos como el robo de datos. Los algoritmos pueden realizar la extracción significativamente más rápido que un ser humano, lo que hace que sea más difícil de detectar y casi imposible de evitar, ya que la máquina copia los datos del perímetro protegido antes de que los defensores puedan reaccionar.

Permitir que otros nodos de la botnet aprendan en forma colectiva e identifiquen las formas de ataque más efectivas

Aprender y compartir información a través de múltiples nodos puede ser una gran ventaja para los atacantes, ya que cada uno de los bots de la red infectada puede probar diferentes técnicas de infiltración e informar los resultados. También puede servirles a los actores maliciosos para aprender más sobre la infraestructura objetivo en menos tiempo.

La IA como parte de los ataques (dirigidos)

Al elegir sus objetivos, los atacantes podrían utilizar la IA para:

Decidir si vale la pena atacar al visitante

Al monitorear el tráfico del sitio Web infectado, el algoritmo puede aprender y seleccionar los visitantes que constituyan los objetivos de ataque más valiosos e infectarlos con el malware.

Identificar una solución de protección específica

El atacante externo puede hacer un reconocimiento de la red objetivo y, en función de las respuestas recibidas (o no recibidas), usar la IA para inferir información sobre las soluciones de seguridad empleadas por la organización objetivo.

La IA como parte de los ataques en entornos móviles

Uso indebido de la popularidad de las aplicaciones móviles

El algoritmo de Machine Learning puede identificar las apps con mods populares y crear su propio mod para modificarlas. Cuando los usuarios incautos descargan estas apps en sus dispositivos móviles, los infectan con malware.

La IA en ataques dirigidos a la IoT

Los dispositivos de la Internet de las cosas, como los routers, las cámaras de seguridad y distintos tipos de controladores, son cada vez más numerosos. En general, las empresas que los utilizan subestiman el hecho de que estos dispositivos son en realidad pequeñas computadoras, y como tales son propensas a presentar vulnerabilidades y ser víctimas de su aprovechamiento por actores maliciosos. Además, los productos de la IoT baratos y mal diseñados a menudo carecen de medidas de seguridad básicas o sus credenciales predeterminadas son débiles: ambas deficiencias pueden permitir la fácil infiltración del malware.

Los atacantes que desean infectar dispositivos de la IoT podrían utilizar la IA para:

- Generar credenciales y usarlas para infiltrarse en otros dispositivos de la IoT similares
- Encontrar nuevas vulnerabilidades en los dispositivos de la IoT
- Si los dispositivos de la IoT son parte de una botnet, el algoritmo puede distribuirse entre todos los nodos para el aprendizaje colectivo
- Conocer los procesos y comportamientos estándar de dispositivos (o grupos) determinados, identificar malware de la competencia y eliminarlo, deshabilitarlo o inutilizarlo

Incluso la IA maliciosa tiene sus limitaciones

Al igual que en cualquier otro campo, la aplicación de la IA en el malware y las actividades maliciosas tiene sus limitaciones. La más importante se documentó en el despliegue del infame Stuxnet, la primera arma cibernética utilizada in the wild.

Esta familia de malware fue muy efectiva para infectar entornos protegidos e incluso aislados por una barrera de aire, lo que le permitió extenderse no solo en los sistemas específicos que quería atacar sino en todo el mundo. Sin embargo, su comportamiento tan agresivo llamó la atención de los investigadores de seguridad, que finalmente identificaron y diseccionaron la amenaza.

Esto también podría aplicarse a futuros ataques basados en la inteligencia artificial. Con el creciente número de infiltraciones, estas amenazas también se volverían más frecuentes y, por lo tanto, visibles, atrayendo más la atención de los defensores de la seguridad, lo que en última instancia conduciría a su detección y mitigación.

LIMITACIONES DEL MACHINE LEARNING

En ESET hemos estado experimentando con varias formas de Machine Learning desde las primeras versiones del producto, y desarrollamos un sistema de detección automatizado que nos ayuda a proteger a nuestros clientes. Sin embargo, este proceso también nos mostró las limitaciones de la tecnología:

Limitación número 1: Conjunto de muestras de entrenamiento

En primer lugar, para utilizar Machine Learning, se necesitan muchas muestras de entrada, cada una de las cuales debe estar correctamente etiquetada. En una aplicación de seguridad cibernética, esto implica tener una gran cantidad de muestras divididas en tres grupos: maliciosas, no infectadas y potencialmente no seguras o no deseadas.

Los investigadores de ESET han pasado más de tres décadas reuniendo y clasificando muestras, y más recientemente eligiendo las muestras que pueden usarse como material de entrenamiento para el motor de ML de ESET, Augur. No obstante, aunque un algoritmo haya sido alimentado con una gran cantidad de datos, no hay garantía de que pueda identificar correctamente todas las muestras nuevas. Por lo tanto, se requiere de la experiencia y verificación humana.

Sin este proceso, incluso una sola entrada incorrecta podría ocasionar un efecto en cadena y llegar a socavar la solución hasta el punto de hacerla fallar. La misma situación se produce si el algoritmo solo utiliza sus propios datos de salida como entradas para seguir entrenándose. En este caso, los errores se consolidan y multiplican, ya que el mismo resultado incorrecto vuelve a ingresar al algoritmo en un bucle y crea más falsos positivos (FP) o pasa por alto los elementos maliciosos.

Otra limitación de las soluciones que solo se basan en el ML o la IA es cuando los ciberdelincuentes deciden atacar una nueva plataforma, como un nuevo lenguaje de script o macro de una aplicación, o un nuevo formato de archivo. En este caso, puede llevar bastante tiempo reunir la cantidad suficiente de muestras "no infectadas" y "maliciosas" para crear un conjunto de entrenamiento.

Limitación número 2: Las matemáticas no pueden resolverlo todo

Algunos fabricantes de seguridad afirman que algunas de estas limitaciones no se aplican a sus algoritmos de ML, ya que pueden identificar cada muestra antes de que se ejecute y determinar si es inofensiva o maliciosa simplemente haciendo cálculos matemáticos. Sin embargo, como lo demostró el famoso matemático, criptoanalista y científico informático Alan Turing (el hombre que rompió el código

Enigma durante la Segunda Guerra Mundial en Bletchley Park en el Reino Unido), este enfoque no es matemáticamente posible.

Ni siquiera una máquina perfecta sería siempre capaz de decidir si una entrada futura desconocida podría llegar a provocar un comportamiento no deseado (en el caso de Turing, sería el de hacer que la máquina reprodujera un bucle indefinidamente). Esto se denomina "problema de la parada" y se aplica a muchos otros campos además del de la informática teórica, donde se originó.

Fred Cohen, el experto informático que definió los virus de computadoras, demostró la aplicación de este principio en la seguridad cibernética mediante otro problema indescifrable: es imposible decir con absoluta certeza si un programa actuará en forma maliciosa si solo se puede analizar durante un tiempo finito. El mismo problema se aplica a futuras entradas o comandos del atacante, que podrían convertir un programa en malicioso.

Por lo tanto, siempre dude cuando un proveedor afirma que su algoritmo de Machine Learning es capaz de etiquetar todas las muestras antes de ejecutarlas (es decir, mediante su ejecución previa) y decidir si son maliciosas o no. Cuando se utiliza un enfoque de este tipo, es necesario bloquear preventivamente una gran cantidad de elementos dudosos, lo que desborda con falsos positivos a los departamentos de seguridad informática de las empresas.

La otra opción sería que utilicen una detección menos agresiva que genere menos falsos positivos; pero si solo se aplica la tecnología de Machine Learning, las tasas de detección se alejarían mucho de la supuesta eficacia infalible del "100%" que estos fabricantes promocionan.

Limitación número 3: Adversario inteligente y adaptable

Otra limitación importante de los algoritmos de Machine Learning en la seguridad cibernética es el adversario inteligente. La experiencia nos enseña que contrarrestar los ataques cibernéticos es como un juego interminable del gato y el ratón. La naturaleza cambiante del entorno de seguridad cibernético hace que sea imposible crear una solución protectora universal capaz de contrarrestar todas las amenazas futuras. El aprendizaje automático no modifica esta conclusión.

Es cierto que las máquinas se han vuelto lo suficientemente inteligentes como para *derrotar a los humanos en el ajedrez* (3) y el *go* (4); sin embargo, estos juegos se basan en reglas estrictas. En cambio, en la seguridad cibernética, los atacantes no siguen pautas ni aceptan limitaciones. Incluso pueden cambiar el campo de juego completo sin ninguna advertencia.

Un buen ejemplo son los automóviles sin conductor. A pesar de la gran inversión en su desarrollo, estas máquinas inteligentes no pueden garantizar el éxito en el tráfico real. Trabajan en áreas limitadas y entornos específicos. Pero imagine qué pasaría si alguien tapara o manipulara las señales de tráfico, o recurriera a actos maliciosos sofisticados, como hacer que los semáforos parpadeen a una velocidad que escapa el reconocimiento humano. Con este tipo de alteraciones en los elementos más críticos del entorno, los autos podrían comenzar a tomar malas decisiones, lo que podría provocar accidentes fatales.

En el ámbito de la seguridad cibernética, la esteganografía es un ejemplo de dicha actividad maliciosa. Los atacantes esconden su código malicioso en archivos inofensivos, como imágenes. Al enterrar el código profundamente entre los píxeles, el archivo (infectado) logra engañar a la máquina, dado que la forma alterada es casi indistinguible de su homólogo no infectado.

Del mismo modo, la fragmentación también puede hacer que una detección basada únicamente en un algoritmo de Machine Learning devuelva una evaluación incorrecta. Los atacantes dividen el malware en partes y las ocultan en varios archivos separados. Cada uno de ellos es inofensivo en forma individual; recién cuando convergen en una endpoint o red comienzan a demostrar un comportamiento malicioso. En tales casos, no hay señales de alarma durante la ejecución previa.

Limitación número 4: Falsos positivos

Se sabe que los cibercriminales trabajan arduamente para evitar la detección y la sofisticación de sus métodos exceden los ejemplos ya mencionados. Usan sus habilidades para ocultar el verdadero propósito de su código "cubriéndolo" mediante la ofuscación o el cifrado. Si el algoritmo no ve más allá de esta máscara, puede tomar una decisión incorrecta. Dejar pasar un elemento malicioso como no infectado o bloquear uno legítimo tiene consecuencias negativas importantes.

Aunque es comprensible por qué un malware no detectado representa un problema para una empresa, es menos obvio con los falsos positivos, que son los errores cometidos cuando una solución de seguridad etiqueta incorrectamente los artículos inofensivos como maliciosos.

No todos los falsos positivos necesariamente conducen al colapso total de la infraestructura de TI de una empresa. Pero algunos de ellos pueden interrumpir la continuidad del negocio y ser potencialmente más destructivos que una infección de malware. Un falso positivo en una fábrica automotriz que etiquetó incorrectamente parte del software de gestión de la línea de producción como malicioso podría interrumpir la producción, provocar demoras masivas, generar un gasto de millones de dólares en daños financieros y afectar la reputación de la empresa.

De todas formas, no es necesario que los falsos positivos rompan procesos críticos para que las organizaciones o su personal de seguridad de TI quieran evitarlos a toda costa. Con decenas o cientos de falsas alarmas diarias (que bien puede ser el caso con una solución de seguridad basada puramente en ML), a los administradores solo les quedan dos opciones:

1. Mantener la configuración estricta y perder días de trabajo solucionando los falsos positivos.
2. Reducir el nivel de protección, lo que puede ocasionar una menor capacidad de detección y crear nuevas vulnerabilidades en la infraestructura de la empresa. Este escenario puede ser provocado y aprovechado fácilmente por un atacante con experiencia cuando la solución de seguridad es demasiado agresiva.

Limitación número 5: El aprendizaje automático solo no es suficiente

Crear defensas efectivas de seguridad cibernética para la red corporativa es similar a proteger una casa. Los propietarios que desean mantener sus hogares seguros deben instalar tantas capas de protección como sea posible, por ejemplo, poner rejas fuertes, cámaras de seguridad, alarmas sonoras y detectores de movimiento en los rincones oscuros.

El enfoque que se debe usar en el entorno corporativo es similar. Sería imprudente confiar en una sola tecnología, por más que sea el último algoritmo de Machine Learning. Con todas sus limitaciones, es necesario usar otras capas protectoras para mantener las endpoints y otros sectores cruciales de la red seguros.

En el entorno de seguridad informática actual, la protección perimetral por sí sola no es suficiente. Se necesitan herramientas adicionales y más elaboradas, como sistemas de detección y respuesta para endpoints (EDR), inteligencia de amenazas, y herramientas que permitan el análisis rápido y confiable de los elementos sospechosos, de modo de proporcionarles a los departamentos de seguridad corporativos los registros y la información forense necesarios.

En consecuencia, si una empresa pretende construir defensas de seguridad cibernética confiables y sólidas, debe seleccionar una variedad equilibrada de soluciones y herramientas que ofrezcan múltiples tecnologías complementarias con altas tasas de detección y un bajo número de falsos positivos. Para volver a la metáfora del hogar: un sistema complejo de seguridad que detecte a los ladrones pero que no haga sonar la alarma cuando el gato del vecino cruce caminando por el jardín.

MACHINE LEARNING SEGÚN ESET: EL CAMINO HACIA AUGUR

En ESET amamos la historia antigua (de hecho, la empresa debe su nombre a una diosa egipcia), por lo que la antigüedad fue, naturalmente, el lugar indicado para buscarle el nombre a nuestro motor de Machine Learning. En la Antigua Roma, un "augur" era un sacerdote que observaba los signos naturales y los interpretaba como indicios de la aprobación o desaprobación divina de una acción propuesta.

No es difícil ver la analogía con la seguridad cibernética, pero a diferencia de los augurios de alquimia en aquel entonces, el motor Augur de ESET basa sus decisiones en la ciencia, las matemáticas y la experiencia previa.

Hay tres tendencias que nos ayudaron a diseñar el motor Augur de ESET:

1. La llegada de los grandes grupos de datos (en inglés "big data") y el hardware más barato
   Gracias a estos cambios, la tecnología de Machine Learning se hizo más asequible, ya sea en el campo de la medicina, la fabricación de automóviles autónomos o las detecciones en seguridad cibernética
2. La popularidad de los algoritmos de Machine Learning
   La cantidad cada vez mayor de aplicaciones basadas en el ML que tienen éxito en la vida real provocó una mayor inversión en este campo, un rápido desarrollo de nuevas funcionalidades, y un incremento en la investigación académica y práctica, lo que contribuye a una mayor disponibilidad de esta tecnología
3. El material de entrenamiento de alta calidad
   Tres décadas de lucha contra los hackers de sombrero negro y sus "productos" le permitieron a ESET construir una "biblioteca de Alejandría" de malware. Este vasto conjunto de muestras muy bien organizadas contiene una recopilación de millones de funcionalidades y genes de ADN de todas las muestras analizadas en nuestros laboratorios de virus. Esta extensa base de datos nos permite crear un conjunto de entrenamiento elegido con precisión, necesario para el desarrollo de Augur

Sin embargo, el auge en estas áreas también trae aparejados nuevos desafíos. Nuestros expertos tuvieron que escoger a mano los algoritmos y enfoques con mejor rendimiento, ya que no todos los algoritmos y tecnologías de Machine Learning se pueden aplicar de la misma forma en entornos de seguridad específicos.

Tras realizar muchas pruebas, decidimos combinar dos metodologías que demostraron ser efectivas hasta el momento:

- El procesamiento con métodos de aprendizaje en profundidad
  Combina unidades de memoria largas a corto plazo (LSTM, por sus siglas en inglés) y redes neuronales totalmente conectadas
- El procesamiento con modelos múltiples (que combina métodos de aprendizaje supervisados)
  Ofrece resultados consolidados de algoritmos de clasificación cuidadosamente seleccionados que integran varios métodos combinados, máquinas vectoriales y árboles de decisión

Esta combinación de algoritmos de clasificación y métodos de aprendizaje en profundidad también aumenta la resistencia de Augur contra la actividad de sus adversarios. Como se documenta en un *paper publicado recientemente* (5) que se centra en el aprendizaje automático de los adversarios, un ataque de evasión que utilice un error genérico o específico para obligar a un motor con una estructura similar a la de Augur a clasificar erróneamente una muestra requeriría una estrategia más compleja.

Cómo Augur procesa las muestras (ver Imagen 9)

El motor de ESET emula el comportamiento de la muestra, y proporciona un conjunto de características y secuencias para su posterior procesamiento. A su vez, ejecuta un análisis detallado de ADN que arroja características numéricas de la muestra. Todos los datos reunidos se combinan en lo que llamamos "xDNA".

En el siguiente paso, Augur analiza la información recopilada utilizando tanto los métodos de aprendizaje en profundidad como el procesamiento con modelos múltiples. Hace poco le incorporamos a Augur un subsistema para clasificar muestras en función de sus binarios. Este subsistema primero desensambla la muestra y usa el resultado para extraer sus características. Luego vectoriza el resultado y lo introduce en la red neuronal.

Cada uno de los subsistemas mencionados en el párrafo anterior produce un valor de probabilidad por separado, que Augur consolida en un valor final mediante el cual etiqueta la muestra como no infectada, potencialmente no deseada/no segura o maliciosa.

Es importante señalar que, a diferencia de los proveedores de seguridad que van más allá de la verdad al promocionar sus productos, ESET utiliza el desempaquetamiento, el análisis de la conducta y la emulación para procesar la muestra. Consideramos que son pasos cruciales para poder extraer adecuadamente las características de una muestra antes de introducirla en el motor Augur. Si solo se utilizaran los datos provenientes del análisis de muestras comprimidas o cifradas, los algoritmos intentarían clasificar también la información irrelevante, lo que generaría resultados en gran medida sin sentido (ver la Imagen 9).

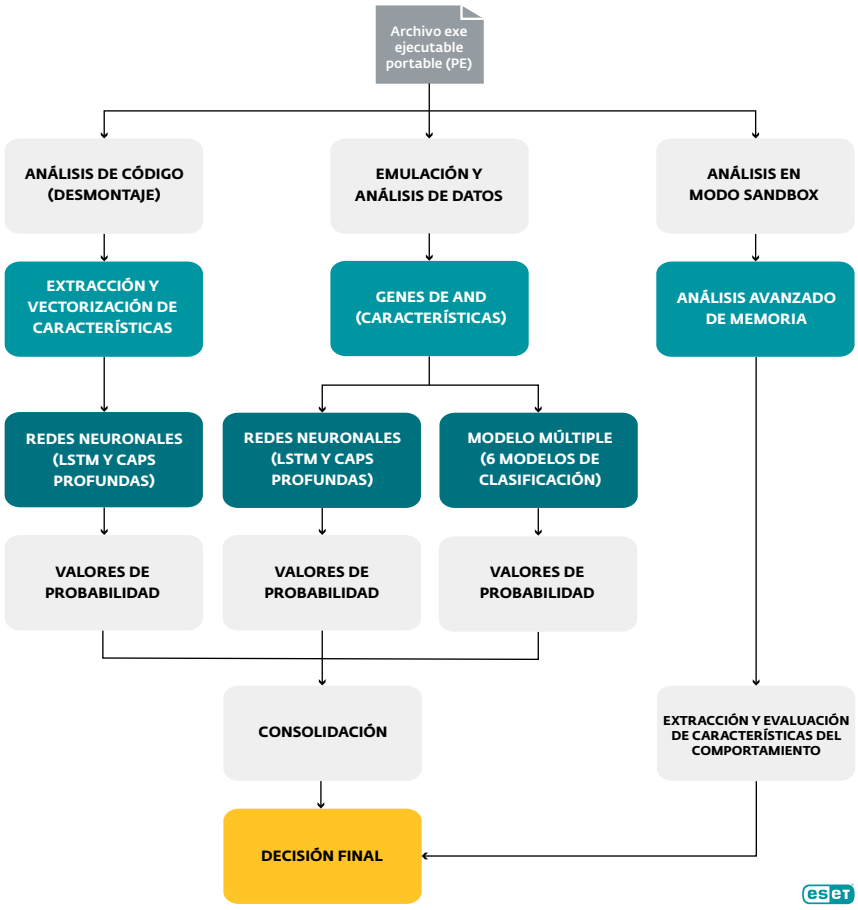

Imagen 9: // Esquema que detalla el funcionamiento del motor de Machine Learning de ESET, "Augur"

Para ofrecer también una perspectiva del mundo real, realizamos una serie de pruebas que demuestran la efectividad del análisis de Augur. Tomamos el antiguo modelo de Augur, que se creó en los primeros meses de 2017, y lo alimentamos con muestras de familias conocidas de malware que causaron estragos en entornos corporativos en los meses subsiguientes: WannaCryptor.D **, Diskcoder.C ((Not) Petya), Diskcoder.D (BadRabbit) y Crysis (una de las familias de ransomware más intrigantes de la actualidad, dirigida a grandes corporaciones y a PYME).

**Para ofrecer también una perspectiva del mundo real, realizamos una serie de pruebas que demuestran la efectividad del análisis de Augur. Tomamos el antiguo modelo de Augur, que se creó en los primeros meses de 2017, y lo alimentamos con muestras de familias conocidas de malware que causaron estragos en entornos corporativos en los meses subsiguientes: WannaCryptor.D **, Diskcoder.C ((Not) Petya), Diskcoder.D (BadRabbit) y Crysis (una de las familias de ransomware más intrigantes de la actualidad, dirigida a grandes corporaciones y a PYME).*

Muestras de malware	Cantidad de muestras	Cantidad de muestras detectadas por Augur	Promedio de detección
Win32_Diskcoder.C	16	10	62.5
Win32_Diskcoder.C (en memoria)	86	85	98.8
Win32_Diskcoder.D	17	14	82.4
Win32_Diskcoder.D (en memoria)	20	20	100
Win32_Filecoder.Crysis	113	112	99.1
Win32_Filecoder.Crysis (en memoria)	30	30	100
Win32_Filecoder.WannaCryptor.D	15	13	86.7
Win32_Filecoder.WannaCryptor (en memoria)	67	67	100

TLos resultados demuestran que, a pesar de que el modelo de Augur utilizado es meses más antiguo que las muestras de malware, la tasa de detección de archivos es bastante alta, en algunos casos incluso perfecta. Sin embargo, el punto más importante para las empresas es que, aunque el archivo se haya ejecutado, Augur fue capaz de identificar correctamente su naturaleza maliciosa en la memoria y les daría a los defensores la oportunidad de detener la amenaza antes de que pudiera causar daños dentro de la infraestructura corporativa. También debemos enfatizar que Augur es solo una capa protectora más implementada en los productos de ESET, y que trabaja con una gran variedad de otras tecnologías que intervienen cuando es necesario.

Augur en los productos de ESET

Los beneficios de Augur ya están disponibles para los clientes de ESET en múltiples frentes. Cada endpoint y dispositivo que tenga habilitado el sistema ESET LiveGrid® utilizará la capacidad de Augur para analizar las amenazas emergentes en una fracción del tiempo que le llevaría a un humano.

Los clientes corporativos de ESET también disfrutarán de los beneficios de Augur a través de dos productos para grandes corporaciones:

1. ESET Enterprise Inspector (EEI) es la herramienta de detección y respuesta para endpoints (EDR, por sus siglas en inglés) de ESET. Recopila datos en tiempo real sobre la actividad que se lleva a cabo en las endpoints y luego los compara con un conjunto de reglas para detectar automáticamente actividades sospechosas. Procesa la información recopilada, la combina y la almacena en un formato que permite la búsqueda, creando una visión general de actividades inusuales y sospechosas. EEI también le proporciona información al equipo de seguridad de la empresa para la investigación forense de incidentes pasados y ofrece capacidades de respuesta para mitigar la presencia de amenazas persistentes avanzadas (APT, del inglés) en la red. Augur está integrado en la exploración de EEI y es esencial para etiquetar actividades y muestras sospechosas.
2. ESET Dynamic Threat Defense (EDTD) es un sistema conformado por múltiples componentes en la nube que le permite a la infraestructura del cliente solicitar información ya analizada de muestras directamente de la base de datos interna de ESET. Si se trata de una muestra nunca antes vista, se sube a los servidores de ESET para su análisis y evaluación detallados a través del motor de ESET (que incluye Augur). Los resultados se devuelven instantáneamente al cliente.

Ambos productos están diseñados para aprovechar las ventajas del motor Augur y trabajar en forma sincronizada con los productos de ESET para endpoints.

CONCLUSIÓN

Como se documenta en este white paper, la tecnología de Machine Learning tiene muchas repercusiones en la seguridad cibernética. Desafortunadamente, también beneficia a los atacantes cibernéticos experimentados, que suponemos que comenzarán a utilizar esta tecnología para proteger su infraestructura maliciosa, mejorar su malware, y encontrar y atacar vulnerabilidades en los sistemas de las empresas.

Debemos destacar que, hasta el momento, no hay evidencia conocida de que se esté usando Machine Learning para "potenciar el malware". Sin embargo, la publicidad exagerada sobre estos temas, sumada al creciente número de noticias sobre fugas masivas de datos y ataques cibernéticos, alimenta los temores de los departamentos de TI corporativos sobre lo que está por venir.

Según los resultados de la encuesta que realizó ESET en los mercados más avanzados del mundo (Estados Unidos, Reino Unido y Alemania), la gran mayoría de los responsables de la toma de decisiones de TI están preocupados por la creciente cantidad y complejidad de los futuros "ataques con inteligencia artificial", así como la mayor dificultad para detectarlos. Como resultado, muchas empresas de seguridad están implementando soluciones que aseguran tener mecanismos de detección avanzados y confiables basados en la "IA".

Estas circunstancias, sumadas a la evolución de las amenazas, son las razones por las cuales los proveedores de seguridad establecidos como ESET mejoran constantemente sus niveles de protección e incorporan Machine Learning (o la IA, en términos más amplios) a sus soluciones. Las pruebas del motor Augur de ESET demuestran lo potente que puede llegar a ser la combinación de una solución en varias capas y el ML, incluso cuando se enfrenta a peligrosas amenazas globales como los ransomware WannaCry, NotPetya, BadRabbit o Crysis.

Con los rápidos avances en el campo de la IA, es difícil predecir cuándo los ataques darán el próximo paso y comenzarán a usar Machine Learning a gran escala, y hasta qué punto las familias de malware se verán potenciadas por las tecnologías descritas como "inteligencia artificial". Pero si en ese momento los usuarios y empresas ya han aplicado las contramedidas y las herramientas de seguridad adecuadas que aprovechan esta tecnología, el impacto y el daño infligido se podrán reducir significativamente.

RESUMEN EJECUTIVO

Como los ataques "potenciados con inteligencia artificial" se están convirtiendo en uno de los temas más solicitados entre los usuarios corporativos, ESET ofrece su punto de vista sobre los posibles usos maliciosos de los algoritmos de Machine Learning (o la IA) sin crear más publicidad exagerada sobre el tema. A su vez, ESET explica las limitaciones que tiene esta tecnología a partir de la experiencia de nuestros equipos de investigación y desarrollo. El paper también incluye los resultados de una encuesta realizada en nombre de ESET a casi 1000 responsables de TI de los Estados Unidos, el Reino Unido y Alemania sobre el uso de la IA y el ML en la seguridad cibernética y las preocupaciones que genera. En la última parte del paper se muestra cómo ESET ha implementado esta tecnología de vanguardia en su motor de múltiples capas y la ha incorporado a su cartera de productos domésticos y corporativos.

HYPERLINKS

1. *http://approximatelycorrect.com/2018/06/05/ai-ml-ai-swirling-nomenclature-slurried-thought/*
2. *http://www.virusradar.com/en/Win32_TrojanDownloader.Swizzor/detail*
3. *https://www.technologyreview.com/s/541276/deep-learning-machine-teaches-itself-chess-in-72-hours-plays-at-international-master/*
4. *http://www.cnbc.com/2017/05/23/googles-alphago-a-i-beats-worlds-number-one-in-ancient-game-of-go.html*
5. *https://arxiv.org/abs/1712.03141*