SOTI

10 YEARS
OF SECURITY INSIGHT

Volumen 10, número 03

Web scrapers en e-commerce

Un peligro para su negocio

Akamai

Estado de Internet en materia de seguridad

Índice

¿Sabía que los bots generan más de la mitad de todo el tráfico web? En concreto, el sector del comercio, debido a su dependencia de aplicaciones y activos web que generan ingresos, se ha visto más afectado por el tráfico de bots de alto riesgo (Figura 1). Y aunque a menudo oímos que los bots están evolucionando, los bots de scraping web son del tipo que está captando la atención de las organizaciones que se apoyan en el e-commerce hoy en día porque sus repercusiones en términos económicos (a menudo ocultas bajo la superficie) difieren de los de otros tipos de bots. La detección de scraper bots también se ha vuelto mucho más difícil debido al aumento de las botnets de inteligencia artificial (IA) y las tecnologías de navegadores sin interfaz, que los hacen extremadamente evasivos. Por ejemplo, uno de los clientes de e-commerce de Akamai tuvo detenido el 99 % del tráfico de alto riesgo sin saber que se debía a bots scrapers.

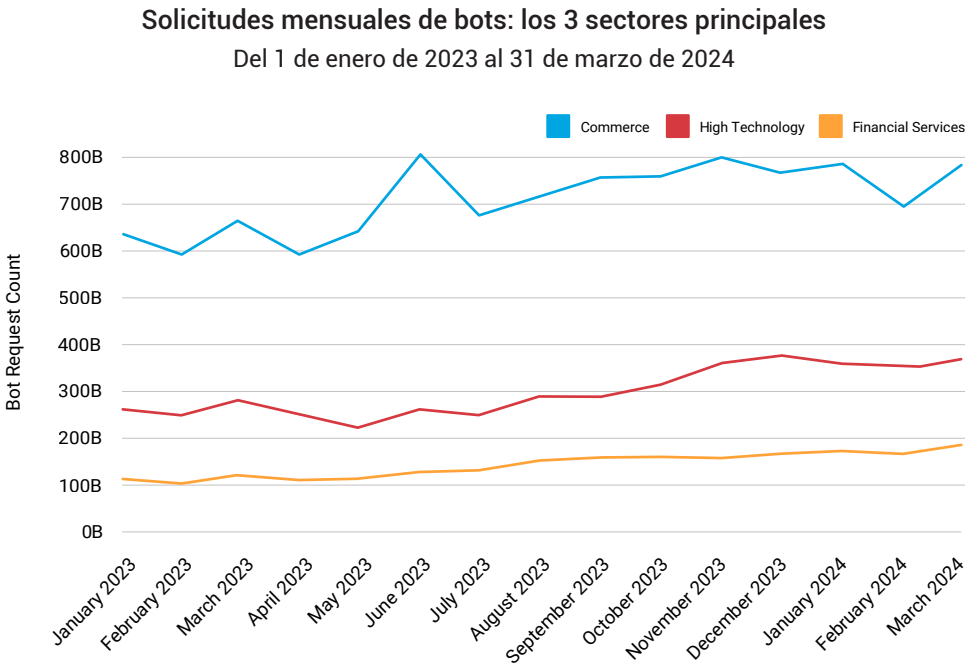

Fig. 1: El comercio es el principal sector por solicitudes de bots y se puede observar un aumento del tráfico global de bots en este sector desde principios de 2023 hasta el primer trimestre de 2024

Por lo tanto, en este informe sobre el estado de Internet (SOTI), nos centramos en la evolución y especialización de estos bots y sus operadores. Aunque los bots existen desde hace tiempo, seguimos viendo su aplicación en una variedad de grupos con el fin de perpetrar ataques criminales, esquemas de fraude y obtener información confidencial de la competencia. Recientemente, hemos observado una tendencia hacia el aumento del uso de todos los bots y un aumento de los impactos negativos de los bots de scraping web en las empresas. Este informe está diseñado para compartir información técnica y metodología de ataque para concienciar sobre este problema creciente en el sector del comercio.

Bots: el bueno, el feo y el malo

Las principales organizaciones centradas en el e-commerce sufren bots que evolucionan continuamente y se especializan en función de lo que pretenden lograr. Dentro del mercado del comercio, hay una gran variedad de tipos de bots que realizan muchas tareas diferentes. Una manera fácil de pensar en ellos es dividirlos en tres grupos: bots buenos, bots malos y bots grises. Los bots buenos ayudan a los clientes a encontrar su sitio. Los bots malos extraen información de su sitio con fines maliciosos. Los bots grises tienden a ser ruidosos, aunque siguen siendo legítimos; en realidad, son una subcategoría de los bots buenos (por ejemplo, los bots de partners que hacen ping constantemente y otras API de programas que realizan llamadas frecuentes).

Por lo tanto, cuando pensamos en chatbots y bots de motores de búsqueda útiles que pueden tener un impacto beneficioso (como responder a las preguntas básicas de los usuarios y proporcionar contenido del sitio web que devuelva resultados de búsqueda más precisos), queremos optimizar esos tipos de bots y, al mismo tiempo, contener los costes de TI. En el caso de los dañinos, como los bots de Credential Stuffing que tratan de obtener acceso no autorizado a la cuenta de un cliente con el fin de apropiarse de dicha cuenta, queremos tomar medidas preventivas sin que ello afecte a la experiencia general del cliente. Uno de los tipos de bots que han aparecido hace poco se está volviendo especialmente problemático, ya que reduce los ingresos, disminuye la fidelidad y aumenta los costes: los bots de scraping web.

Los bots de scraping web, una botnet utilizada para extraer directamente datos y contenido de sitios web en Internet, son únicos. Exigen atención debido a que operan de forma diferente; además, su impacto en la empresa y los mecanismos de detección varían con respecto a los de otros bots. Los scrapers web también son polifacéticos, ya que sus casos de uso varían en función de cómo las organizaciones y los operadores monetizan la información que recopilan estos bots. Independientemente del objetivo concreto, los scrapers están afectando a los ingresos, aumentando los costes de TI y perturbando la experiencia general del cliente.

En este informe SOTI, examinamos el impacto del scraping en el e-commerce y analizamos por qué los propietarios de las empresas (pensando en lo digital, el marketing, la marca, las finanzas, el riesgo y la seguridad) deben tener un interés común en detener a los scrapers abusivos. Para comprender mejor estos impactos, es fundamental ver la imagen completa de por qué han evolucionado los bots de scraping web, para qué se utilizan, cómo funcionan, de qué manera afectan y qué pueden hacer las empresas al respecto.

Información clave del informe

El scraping web no representa solo una cuestión de fraude o un problema de seguridad, también es un problema empresarial. Los scraper bots tienen un efecto negativo en muchas facetas de la organización, como los ingresos, la ventaja competitiva, la identidad de marca, la experiencia del cliente, los costes de infraestructura y la experiencia digital, por nombrar algunos.

Según un caso real de investigación de Akamai, el 42,1 % de la actividad general de tráfico procedía de bots, y el 65,3 % de ese tráfico de bots procedía de bots maliciosos. Y un total del 63,1 % del tráfico de bots maliciosos utilizaba técnicas avanzadas.

La tecnología de navegadores sin interfaz ha cambiado el panorama de los scrapers, lo que requiere un enfoque para gestionar este tipo de actividad de bots más sofisticado que otras medidas de mitigación basadas en JavaScript.

Entre las consecuencias técnicas a las que deben hacer frente las empresas tras un ataque de scraping, independientemente de si las intenciones eran buenas o malas, encontramos la ralentización del rendimiento y el falseamiento de los datos de la web, las consecuencias del robo de credenciales, el aumento de los costes informáticos, etc.

Es importante observar y comprender los diferentes patrones de tráfico para identificar si un sitio web está incurriendo en tráfico humano, de bots básicos o de bots sofisticados. Estos patrones pueden variar de circadianos a intermitentes y continuos.

Bots buenos frente a bots malos

Empecemos por lo básico: un bot (abreviatura de "robot"), es un programa informático que puede realizar tareas automatizadas de forma más rápida y precisa que una persona. Los distintos roles y tipos de bots se dividen en dos categorías principales: bots buenos y bots malos (Figura 2). Los bots grises son una subcategoría de los bots buenos, pero los combinaremos con los bots buenos por ahora para simplificar la comparación.

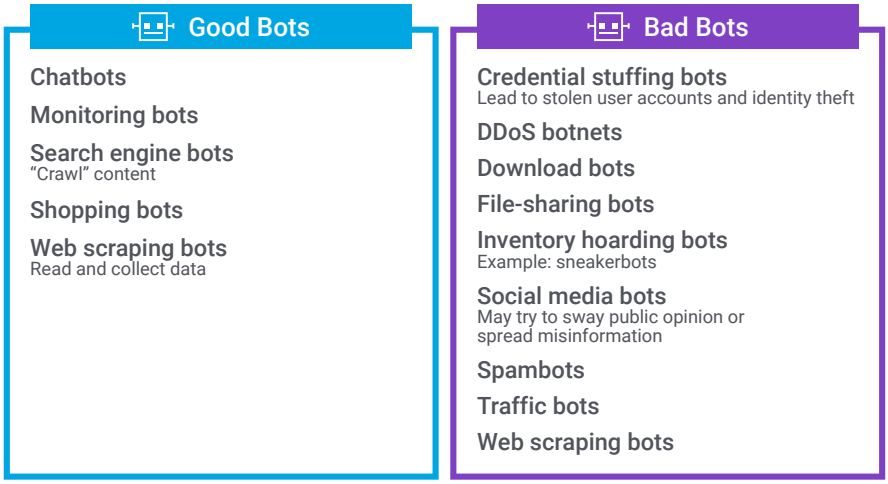

Fig. 2: Una comparación en paralelo, con ejemplos, de bots buenos y bots malos

Los bots buenos son bots útiles que ayudan a proporcionar herramientas y servicios, mientras que los ciberdelincuentes y los estafadores suelen utilizar los bots malos con intenciones maliciosas. Un ejemplo de este tipo de malicia es un bot de tráfico que imita el comportamiento humano online para aumentar los clics y el tráfico en un sitio web (es decir, cometer fraudes mediante anuncios).

Los bots de scraping web aparecen en las categorías de bots buenos y malos. La distinción tiene que ver con el modo en que las organizaciones utilizan la información que recopilan estos bots. Ahora nos centraremos más en varios casos de uso relacionados con los efectos positivos y negativos de los scraper bots a los que se enfrentan algunos de los principales retailers y marcas de e-commerce del mundo.

Introducción al scraping

Las empresas de e-commerce suelen utilizar el scraping web. En los sectores del turismo y la hostelería, por ejemplo, los agregadores de viajes extraen contenido dinámico de sus hoteles y aerolíneas asociados para mantenerse al día sobre la disponibilidad y los precios. Este tipo de scraping se prevé y las empresas utilizan controles de bots comunes para acelerar los scrapers en momentos del día en los que los usuarios reales desean realizar una reserva. Las organizaciones también utilizan proveedores de servicios de extracción de datos para recopilar clientes potenciales y otra información relacionada de los competidores. Además, los bots de scraping se pueden utilizar para analizar datos e identificar tendencias.
El scraping también puede ser beneficioso para la revisión del sitio para mejorar las ofertas y los servicios online, así como para permitir a los consumidores potenciales encontrar más fácilmente los productos de la empresa, como a través de un motor de búsqueda. Todas estas acciones pueden ayudar a las empresas a lograr una ventaja competitiva. Sin embargo, no se puede negar que muchas entidades están utilizando scrapers por razones menos encomiables.

El scraping cambia y los clientes se dan cuenta

Lamentablemente, a menudo oímos hablar de consumidores que han sido víctimas de estafas de phishing. En este caso, los bots scrapers se pueden haber utilizado para obtener imágenes de productos, descripciones e información sobre los precios para crear sitios de phishing o tiendas falsas para sustraer la información de tarjetas de crédito o las credenciales de los usuarios. Estos sitios falsos o de phishing son una forma de suplantación de la marca, en la que la propiedad intelectual de las organizaciones víctimas se utiliza para establecer confianza con los clientes potenciales.

Algunas de las marcas de e-commerce más grandes del mundo se han visto afectadas por sitios falsos, campañas de phishing y el robo de datos web de la empresa como parte de campañas de suplantación de marca (Figura 3). Lamentablemente, cuando los sitios de phishing tienen éxito, las marcas legítimas se enfrentan a las consecuencias de la pérdida de confianza y fidelidad de los clientes.

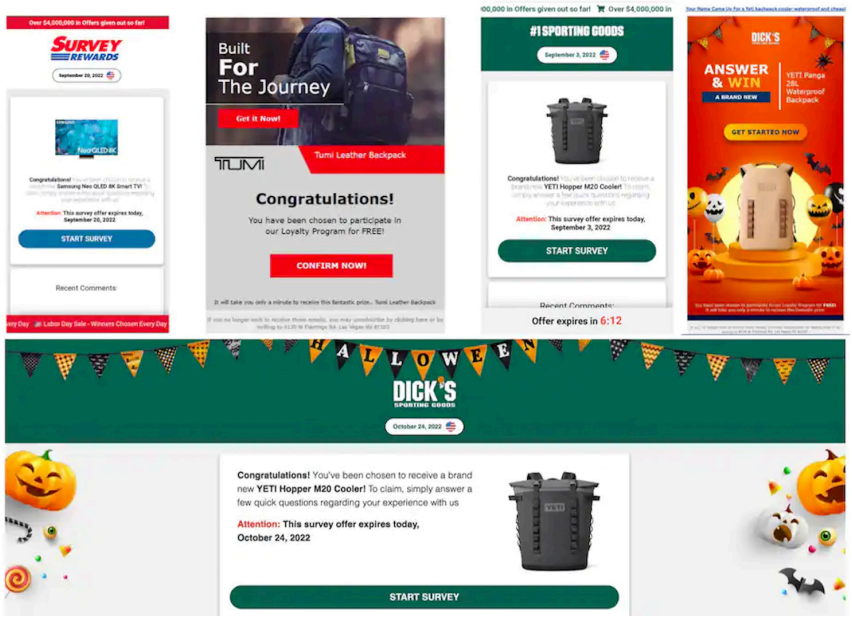

Fig. 3: Un ejemplo de algunas de las principales empresas de e-commerce que han sido víctimas de la suplantación de la marca

La especulación también se puede atribuir al scraping web, ya que los especuladores pueden extraer información de un sitio en busca de productos disponibles y comprarlos antes de que los clientes legítimos tengan la oportunidad de hacerlo (Figura 4).

Fig. 4: Casos de uso de scrapers

Los atacantes que llevan a cabo este tipo de actividades de scraping perjudiciales son conscientes de los efectos que sus objetivos maliciosos tienen en las víctimas. Esto incluye los impactos negativos de la inteligencia y el espionaje de la competencia, la acaparación de inventario/reventa, la falsificación y la suplantación de sitios web/bienes, y el scraping de sitios del sector multimedia y la reutilización del contenido (Tabla 1). Y no existen leyes que prohíban explícitamente el uso de bots scrapers.

Impacto	Descripción
Espionaje e información sobre la competencia	Los competidores utilizan la información del sitio de una organización para reducir los precios, realizar cambios en sus ofertas y hacerse una idea de las nuevas oportunidades y amenazas.
Acaparamiento/ extracción de inventario	Los especuladores hacen ping de manera constante a los sitios objetivo en busca de productos que estén disponibles y los añaden a los carritos, con lo que esos productos dejan de estar disponibles para los clientes reales.
Falsificación de artículos y suplantación de sitios web	Los falsificadores utilizan el contenido de los scrapers para crear sitios falsos y catálogos de productos para engañar a los clientes y hacerles creer que están comprando productos legítimos en lugar de falsificaciones.
Scraping de sitios del sector multimedia y reutilización del contenido	Los atacantes pueden extraer artículos de noticias, blogs y otro contenido y colocarlo en sus propios sitios, lo que provoca que la organización original pierda visitantes y posibles ingresos por publicidad. Las tarifas de publicidad a menudo se basan en el número de visitantes/público del sitio, por lo que menos visitantes significa que el sitio del sector multimedia pierde ingresos que habría obtenido gracias a unas tarifas publicitarias más altas.

Tabla 1: Impactos negativos intencionados causados por los scrapers web

Los efectos secundarios generales del scraping web

Independientemente de la intención del scraping web, las organizaciones tienen que hacer frente a los gastos derivados de sus efectos secundarios. Algunas empresas pagan por servicios de scraping beneficiosos, pero las empresas que sufren ataques de scraping están incurriendo en sus propios costes. Entre ellos se incluyen los gastos de las soluciones antibots y los impactos económicos negativos de la degradación del rendimiento del sitio y la contaminación de las métricas clave (Tabla 2).

Impacto	Descripción
Aumento de los costes de servidores, CDN y nube para atender el tráfico de bots	Esto afecta a los ingresos y provoca pérdidas de reputación por el uso del contenido por parte de competidores, atacantes y falsificadores.
Degradación del rendimiento del sitio	Puesto que los bots scrapers se ejecutan de forma continua hasta que se detienen, aumentan los costes de servidores y distribución, ya que las organizaciones deben gestionar el tráfico de bots no deseado. Además, se degrada la experiencia de usuario debido aun rendimiento más lento de los sitios y las aplicaciones.
Contaminación de las métricas clave	La actividad de bots no detectada distorsiona en gran medida las métricas clave, como la tasa de conversión del sitio, en las que confían los equipos empresariales para tomar decisiones de inversión, como las estrategias de posicionamiento de productos y las campañas de marketing.

Tabla 2: Impactos negativos no intencionados causados por los scrapers web

Scraping de alquiler: servicios de scraping web de terceros

Como hemos mencionado, los bots de scraping web se pueden utilizar con fines buenos o malos. A diferencia de los bots utilizados para los ataques de Credential Stuffing, que son bots maliciosos conocidos y, por lo tanto, se bloquean con razón, hay empresas que ofrecen bots de scraping web legítimos. Muchas organizaciones utilizan estos servicios de scraping web de terceros para extraer y proporcionar datos a su propia organización, lo que puede ser beneficioso, especialmente en el mundo del marketing competitivo.

Decenas de estas empresas proporcionan diferentes tipos de scraping web/servicios de extracción de datos; hay incluso conferencias que los promueven. Por ejemplo, Bright Data organiza una conferencia denominada ScrapeCon que reúne a expertos en evitar la detección de bots para que las empresas puedan aprender a extraer datos. La Tabla 3 incluye ejemplos de los niveles de servicios que pueden proporcionar empresas de scraping web de terceros.

Nivel de servicio 1	Los servicios de proxy pueden formar parte del scraping y ofrecer una infraestructura que podría incluir las direcciones IP móviles y residenciales de los centros de datos.
Nivel de servicio 2	Este segundo nivel también puede incluir la extracción automatizada de datos que limpia y estructura los datos para facilitar su uso por parte de los miembros del equipo de ciencia de datos del cliente, que extraen información valiosa para orientar las decisiones de la empresa.
Nivel de servicio 3	El nivel más alto puede añadir la extracción de la información empresarial real en sí, lo que puede mejorar aún más el proceso de toma de decisiones para las empresas. Se denominan "botnets de IA".

Tabla 3: Varios niveles de servicios proporcionados por empresas de scraping web de terceros

Los clientes pueden elegir cualquiera de estos niveles de servicio, desde el más básico hasta el más avanzado, así como la frecuencia de la recopilación de datos, y pueden especificar sus objetivos. A menudo, el nivel de servicio prestado, o botnet elegida, depende del nivel de protección que necesitan superar. Una botnet más básica puede recopilar datos a través de un script avanzado con unos miles de servidores proxy ubicados en centros de datos que equilibran la carga de tráfico. Si la protección es lo suficientemente rudimentaria, la botnet podría utilizar esta técnica para pasar por las defensas de gestión de bots y el firewall de aplicaciones web de la infraestructura de seguridad.

Sin embargo, si la protección es más avanzada, puede ser necesario un enfoque más sofisticado del scraping, como un ataque de navegador sin interfaz. Esto se aplica independientemente de si el scraping lo realiza un agente con buena o mala intención. Y no es barato, ya que las empresas van a incurrir en costes que generalmente son mucho más altos para la infraestructura más sofisticada que para el nivel de servicio básico. Una defensa avanzada puede incluir tecnologías de desafío (como CAPTCHA o prueba de trabajo), varias capas de detección diseñadas para la evaluación de huellas dactilares en el lado del cliente y un análisis de las características del protocolo de transferencia de hipertexto (HTTP) y de la seguridad de la capa de transporte (TLS).

El proceso de scraping de las botnets de IA

Aunque los scrapers web básicos pueden ser más coherentes en sus técnicas de scraping, las botnets de IA tienen la capacidad de detectar y extraer datos y contenido no estructurados en un formato o ubicación menos coherente. Además, las botnets de IA pueden utilizar la inteligencia empresarial real para mejorar el proceso de toma de decisiones. Las sofisticadas botnets de IA, mencionadas en la Tabla 3, nivel de servicio 3, siguen un proceso de tres pasos para extraer datos. Funcionan recopilando, extrayendo y procesando datos (Figura 5).

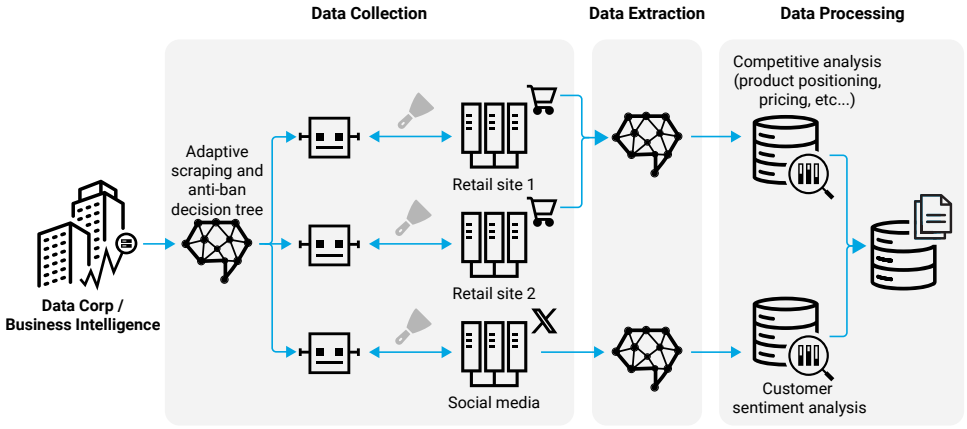

Fig. 5: Representación de una botnet de IA y su proceso de tres pasos

Examinemos estos tres pasos con más profundidad para comprender mejor lo que implican.

Recopilación de datos

El scraping web implica la organización de los datos que se extraen de un sitio web, o sitios web, de modo que las organizaciones puedan producir nuevos conjuntos de datos que se puedan aplicar y analizar como consideren oportuno. Y comienza con la recopilación de datos.

Es necesario que la recopilación de datos conste de un scraping adaptable combinado con tecnologías "anti-ban" o de "detección antibots" para funcionar de forma rápida y fluida. Estas tecnologías se establecen como árboles de decisión para detectar diversos aspectos de cualquier protección que pueda haber aplicada. La resiliencia es el quid de la cuestión aquí. La protección contra bots puede incluir la huella dactilar de JavaScript, la huella dactilar de HTTP y TLS (evaluación de los encabezados HTTP y el protocolo de negociación TLS) y la detección de reputación del protocolo de Internet (IP) (Figura 6). Algunos de estos flujos de trabajo pueden incluir el aprendizaje automático (ML), especialmente al recopilar estadísticas sobre el índice de éxito; ajustar la estrategia de cookies, el encabezado HTTP y los parámetros de TLS; y evaluar el código de huella dactilar de JavaScript. Aquí es también donde un navegador sin interfaz puede entrar en juego.

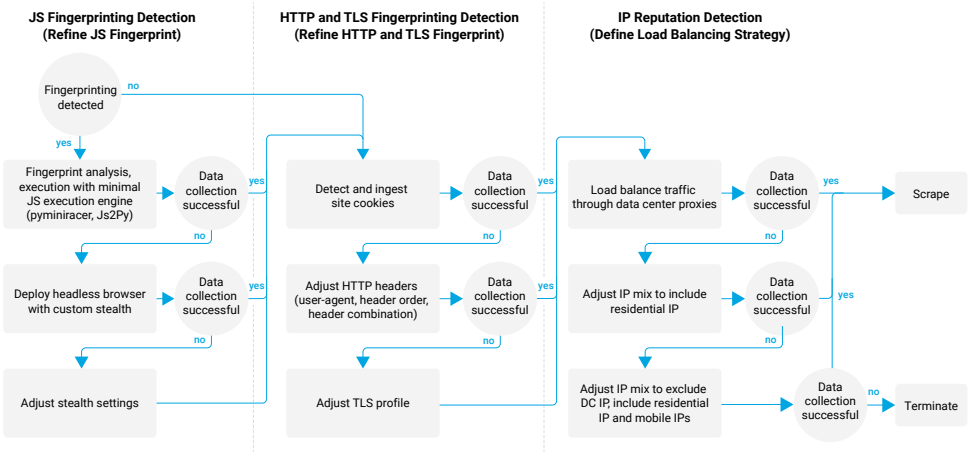

Fig. 6: Al tratar de recopilar datos, este árbol de decisiones de detección antibots intenta evitar las huellas digitales de JavaScript, las huellas digitales de HTTP y TLS y la detección de reputación del protocolo de Internet (IP)

El navegador sin interfaz

Un navegador sin interfaz es un navegador web que no tiene una interfaz gráfica de usuario (GUI). Esto significa que las personas no pueden interactuar directamente con la página web en la que aparece el navegador sin interfaz y el navegador se ejecuta a través de una interfaz de línea de comandos (CLI) o mediante una comunicación de red. En el caso de Selenium, un conocido navegador sin interfaz de código abierto, el navegador está automatizado y se utiliza ampliamente para el scraping web. Esto puede ser muy útil para los buscadores de datos que están intentando extraer contenido dinámico.

Los navegadores sin interfaz también pueden permitir que las capturas de pantalla y el código del sitio web se copien de forma eficaz y que los datos elegidos se extraigan sin mostrar toda la página. Sin embargo, los ataques a través de navegadores sin interfaz resultan caros y a veces se pueden detectar mediante las huellas digitales que dejan. Sin embargo, los gastos de otra infraestructura sofisticada son similares a los de los navegadores sin interfaz, es decir, que generalmente son altos.

Extracción y procesamiento de datos

La información extraída normalmente consta de contenido HTML y JSON. De todos los datos extraídos, solo una parte de ellos puede ser útil para el análisis. Por ejemplo, el análisis de la competencia suele incluir precios, descuentos, inventario y números SKU de productos, categorías y descripciones. Es posible extraer información esencial de forma automática mediante modelos de aprendizaje automático que se pueden formar con varias estructuras y formatos de datos para reconocerla. Esto permite evitar todo el trabajo de procesamiento adicional que debe realizarse para extraer manualmente los datos, así como la necesidad de estudiar la estructura del código del contenido HTML y JSON. Además, la estructura del código del contenido puede cambiar a medida que el diseño del sitio evoluciona. También es necesaria una lógica de aprendizaje automático adicional para el procesamiento si hay varios sitios web implicados como parte del alcance del análisis.

Caso real: ventajas de las soluciones de detección de scraping web

Los investigadores de Akamai observaron a un subconjunto de clientes de e-commerce que estaban protegidos por una solución de scraping web que detectaba actividades de scraping, y examinaron el desglose de la actividad de tráfico durante una semana. Esto ascendió a un tamaño de muestra de aproximadamente 6900 millones de solicitudes. El análisis solo tuvo en cuenta las solicitudes HTML y AJAX. El contenido estático (imágenes, JavaScript, hojas de estilo) no se incluyó en el análisis, ya que la mayoría de los bots no solicitan contenido estático; esta omisión también ayudó a evitar inflar innecesariamente los datos.

Akamai Content Protector clasificó la actividad general, que consistió en un 49,3 % de tráfico humano de bajo riesgo, un 42,1 % de tráfico de bots (un 27,5 % de bots maliciosos de alto riesgo y un 14,6 % de bots buenos) y un 8,7 % de tráfico no clasificado de riesgo medio (Figura 7).

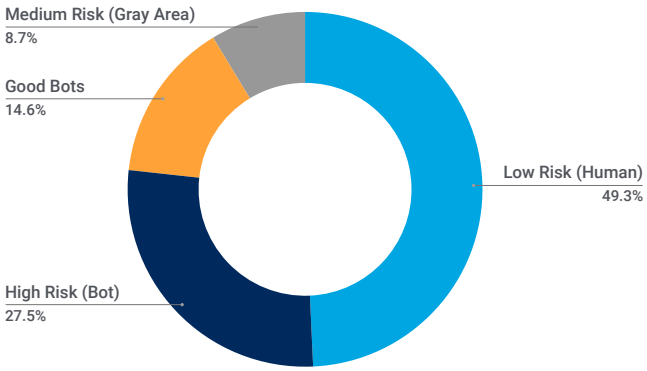

Fig. 7: Desglose de la clasificación de la actividad de tráfico

La Figura 8 muestra que del 42,1 % del tráfico procedente de bots, el 65,3 % procedía de scrapers considerados bots maliciosos y el 34,7 % restante procedía de scrapers clasificados como bots buenos (por ejemplo, motores de búsqueda web, SEO, redes sociales y publicidad online).

Fig. 8: Tráfico de bots buenos frente a tráfico de bots malos

También se midieron los niveles de sofisticación de los bots malos de alto riesgo que contribuyeron al 65,3 % del tráfico general de bots. El 37 % de ese tráfico procedía de botnets con scripts básicos que son fáciles de detectar mediante métodos sin estado sencillos, el 47,6 % procedía de botnets con scripts más avanzados que requieren métodos de detección con estado más avanzados mediante el aprendizaje automático y el 15,5 % procedía de navegadores sin interfaz que requieren métodos avanzados de detección de huellas digitales de JavaScript y con estado (Figura 9).

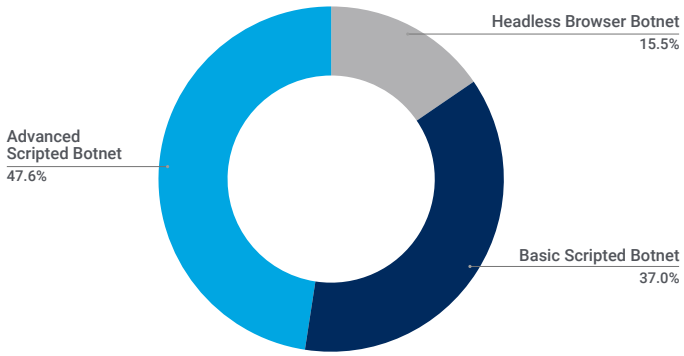

Fig. 9: Distribución del tráfico de bots malos en función de su sofisticación (los totales no suman el 100 % debido al redondeo)

A partir de estos datos, podemos ver que los scrapers de bots malos son considerablemente más numerosos que los scrapers de bots buenos y que casi la mitad del tráfico total estaba compuesto por bots, y las botnets con scripts avanzados producían el mayor tráfico de bots malos (47,6 %).

El sitio web funcionará de forma mucho más rápida y eficiente, y las métricas del sitio se leerán con mayor claridad, una vez que se hayan implementado las defensas contra estos bots y se hayan eliminado los scrapers. Y estos resultados se traducirán en mejores experiencias de usuario/cliente. Como se muestra en la Figura 10, el número de solicitudes de bots de alto riesgo disminuyó sustancialmente una vez activada la mitigación.

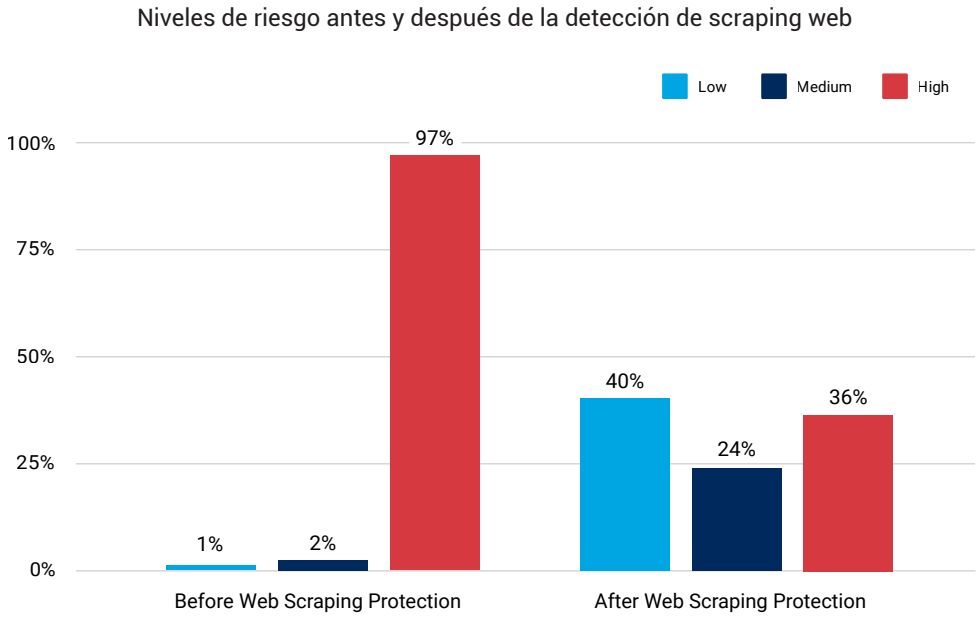

Fig. 10: Niveles de riesgo antes y después de la mitigación con Content Protector

Protección y mitigación

En esta sección se proporcionan algunos indicadores fundamentales para la detección de scrapers web e información sobre herramientas que pueden proporcionar medidas defensivas contra ellos.

Detección de scrapers básicos

Aunque los scrapers sofisticados pueden ser difíciles de detectar, las soluciones de gestión de bots pueden evitar la recopilación de datos por parte de todo tipo de scrapers intrusivos y, en concreto, pueden buscar las siguientes características para detectar los bots de scraping web más sencillos:

- Solicitudes procedentes de navegadores y versiones de sistemas operativos más antiguos
- Anomalías en la firma del encabezado HTTP
- El uso de versiones antiguas de HTTP (por ejemplo, v1.1) en lugar del HTTP v2 más común o el HTTP v3 emergente
- Solicitudes procedentes de miles de centros de datos/servicios en la nube

Detección de scrapers más avanzados

Ninguna de las características de la lista anterior se podrá observar para los scrapers más avanzados. Estas son algunas de las características de los scrapers más sofisticados:

- Solicitudes procedentes de la versión más reciente del navegador y del sistema operativo
- El conjunto de encabezados HTTP parece idéntico al del navegador legítimo
- El uso de HTTP v2
- Solicitudes procedentes de cientos de miles de direcciones IP residenciales y móviles

Identificación de patrones de tráfico

Existen algunos indicadores clave que pueden identificar si el tipo de tráfico presente en un sitio web es humano (Figura 11), bot básico (Figura 12) o bot sofisticado (Figura 13).

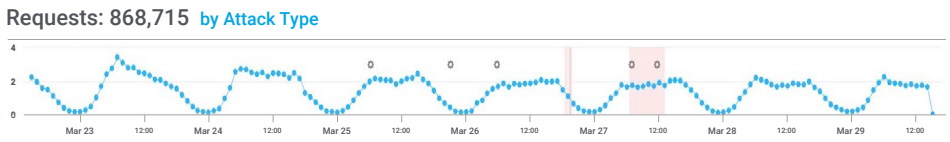

Fig. 11: El tráfico de usuarios legítimos suele mostrar un ciclo circadiano de actividad

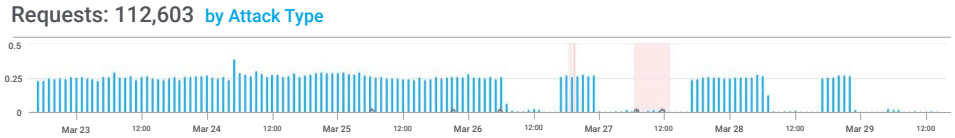

Fig. 12: El tráfico típico de bots muestra una actividad normal con interrupciones ocasionales

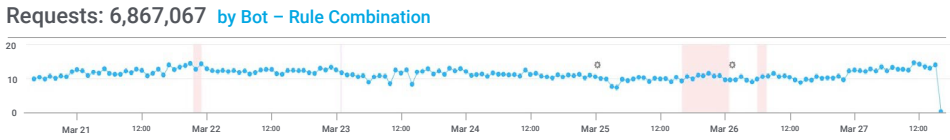

Fig. 13: Los bots más sofisticados muestran el tráfico continuamente, de día y de noche

A menudo también vemos botnets que se encuentran en medio, con una estrategia de balanceo de carga débil, pero una estrategia de huella digital sofisticada (o viceversa). Sin embargo, las botnets más avanzadas pueden ser tan sofisticadas que puede parecer que tienen una huella digital perfecta o incluso reproducen un patrón de tráfico de usuario legítimo.

Además de estar al tanto de estos bots de scraping, las herramientas que protegen contra el scraping web, como un protector de contenido, pueden ofrecer ventajas especiales y una navegación más fluida en aguas agitadas llenas de scrapers. Entre las ventajas se incluyen:

- Mayores tasas de conversión y menores costes de TI
- Métricas más precisas, que pueden conducir a mejores decisiones de inversión y a un aumento de los ingresos
- Reducción de la presión sobre los precios, lo que puede traducirse en ventas que se salvan de la rebaja por efecto de la competencia
- Clientes satisfechos que pueden acceder a los productos deseados y aumento de los ingresos gracias a las oportunidades de venta incremental que surgen cuando los consumidores añaden productos adicionales a sus carritos una vez que se han hecho con el artículo que buscaban
- Preservación de la reputación de la marca, ya que se protege a los clientes de falsificaciones de mala calidad que creen que son bienes legítimos del vendedor original
- Retención de los ingresos por productos y fidelización de los clientes
- Aumento y protección de los ingresos por publicidad
- Retención del público y los visitantes del sitio

Consideraciones sobre cumplimiento

La norma de seguridad de datos del sector de las tarjetas de pago (PCI DSS) v4.0 está en vigor, y muchos de los cambios que incluye se deben a tendencias de amenazas que siguen afectando a las empresas. La visibilidad es un elemento clave para abordar estos ataques. Tanto si se encuentran en su entorno JavaScript histórico como en las API utilizadas para facilitar la transformación, es fundamental detectar y corregir rápidamente estos ataques.

También observamos tendencias emergentes en lo que respecta al cumplimiento en el nuevo marco de ciberseguridad del Instituto Nacional de Normas y Tecnología (NIST) versión 2.0, que ha añadido una función de gobierno. El NIST representa a menudo la base de una serie de normativas gubernamentales y se integra en muchos marcos de ciberseguridad comercial. Por lo tanto, este es un buen momento para revisar la nueva guía y utilizarla a la hora de actualizar sus políticas o analizar su documentación actual y ver así dónde presenta lagunas.

Para las empresas que cotizan en bolsa y aquellas que utilizan principios de contabilidad generalmente aceptados (GAAP), otra área de cumplimiento es la materialidad de la ciberseguridad. La necesidad de definir riesgos y amenazas materiales requiere la colaboración entre el equipo de dirección. Una vez que identifique las amenazas materiales (como el ransomware), debe asignar mecanismos de mitigación (como la microsegmentación). Asegúrese de que sus planes de gestión de crisis cumplan con los plazos de divulgación y de contar con una guía para el peor de los casos, para el que tendría que presentar un Formulario 8-K sobre incidentes de ciberseguridad de la Comisión de Bolsa y Valores (SEC).

Conclusión

Esperamos que este informe le proporcione información sobre un área que podría estar teniendo un impacto económico negativo en su organización. Los bots están afectando a sus sitios en un volumen cada vez mayor y es importante optimizar los bots beneficiosos, mitigar los bots maliciosos y garantizar una experiencia de usuario sin problemas. Se trata de un problema de seguridad que afecta a la empresa. Al igual que con todos los problemas de seguridad, el primer paso es obtener visibilidad, el segundo paso es analizar el impacto y el último paso es determinar el retorno de la inversión para el riesgo y los ingresos, de modo que pueda implementar los controles de seguridad adecuados.

No se puede proteger lo que no se puede ver, por lo que ahora es el momento de determinar dónde hay deficiencias de visibilidad. Para ello, debe determinar el nivel de actividad de scraping web en sus sitios y su intención. El panorama de bots está formado por bots buenos y bots malos, y los bots scrapers se encuentran en ambas categorías, en función de su uso. Aunque la línea entre los bots scrapers beneficiosos y perjudiciales puede ser borrosa, la evolución de la sofisticación de los bots (por ejemplo, los scrapers web que llevan a cabo ataques de navegador sin interfaz) continúa. Todo esto va acompañado del enorme impacto que tienen los bots de scraping web entre las entidades de e-commerce, tanto en los costes de TI como en la experiencia de cliente. Es fundamental asegurarse de que dispone de las herramientas necesarias para analizar la actividad de los bots y el impacto en su sitio.

Lo que no desea es que los atacantes ejecuten su modelo de negocio delictivo en sus sitios web y cometan toda una serie de actividades maliciosas, como el cobro de puntos de fidelidad, la realización de pedidos fraudulentos o incluso la comisión de fraudes en devoluciones. Tampoco quiere que le afecten los bots que hacen compra masiva de entradas para eventos con plazas limitadas o los bots que acaparan la compra de artículos de moda. Los bots se pueden utilizar para facilitar la apertura abusiva de nuevas cuentas aprovechando ofertas especiales, lo que afecta al análisis y los costes de las campañas. Las grandes botnets que ejecutan ataques distribuidos de denegación de servicio (DDoS) pueden saturar las aplicaciones web y provocar una mala experiencia de usuario o la incapacidad de realizar pedidos o reservas, lo que provoca pérdidas de ingresos y problemas con los clientes. Los bots pueden incluso imitar el comportamiento humano online para aumentar los clics y el tráfico en un sitio web, lo que sesga tanto el análisis de marketing como el de rendimiento de experiencias digitales cuidadosamente diseñadas. Definitivamente, no desea nada de eso.

Como hemos señalado anteriormente, más de la mitad del tráfico web de comercio mundial está compuesto por bots y los niveles de tráfico de bots siguen aumentando. Akamai ha basado la información y los consejos de este informe en nuestra plataforma de seguridad, que incluye protección de contenido con defensa contra el scraping web. Colaboramos con muchos líderes de e-commerce, por lo que queríamos compartir medidas de protección y mitigación que las empresas puedan utilizar para proteger mejor a sus clientes. Prevemos un aumento en el uso, las opciones de nivel de servicio y los tipos de bots de scraping web disponibles. Por lo tanto, es necesario evaluar continuamente la estrategia de riesgo de su empresa y determinar si sus controles de seguridad actuales satisfacen el apetito por el riesgo de su equipo de dirección.

Manténgase informado sobre nuestra investigación más reciente visitando nuestro Centro de investigación sobre seguridad.

Metodologías

Datos de Content Protector

Esta muestra de datos describe las clasificaciones de nivel de riesgo que nuestra herramienta Content Protector asigna al tráfico que supervisa. Estas clasificaciones se utilizan para detectar tanto las actividades de scraping de bots como para determinar si se trata de un bot bueno o malo. Dado que la mayoría de los bots no solicitan contenido estático, este análisis solo tuvo en cuenta las solicitudes HTML y AJAX para evitar inflar innecesariamente los datos.

Esta muestra de datos abarcó el periodo de una semana comprendido entre el 12 y el 19 de abril de 2024. El tamaño total de nuestra muestra consistía en más de 6500 millones de solicitudes.

Ataques de bots

Estos datos describen las alertas en la capa de aplicación sobre el tráfico observado a través de nuestro firewall de aplicaciones web (WAF) y la herramienta de gestión de bots. Las alertas de bots se activan cuando detectamos una carga de bots en una solicitud a un sitio web, una aplicación o una API protegidos. Estas alertas de bots las pueden activar tanto bots maliciosos como bots legítimos. Las alertas no indican que un ataque haya conseguido su objetivo. Aunque estos productos permiten un alto nivel de personalización, recopilamos los datos presentados aquí de una manera que no tiene en cuenta las configuraciones personalizadas de las propiedades protegidas. Los datos se extrajeron de una herramienta interna de análisis de eventos de seguridad detectados en Akamai Connected Cloud, una red de aproximadamente 340 000 servidores repartidos entre más de 4000 centros, casi 1300 redes y más de 130 países. Nuestros equipos de seguridad utilizan estos datos, medidos en petabytes mensuales, para investigar ataques, detectar comportamientos maliciosos y proporcionar información adicional a las soluciones de Akamai.

Estos datos cubren el periodo de 15 meses que abarca desde el 1 de enero de 2023 hasta el 31 de marzo de 2024.

Créditos

Editor jefe

Lance Rhodes

Editorial y redacción

David Senecal
Maria Vlasak

Revisión y expertos en la materia

Mitch Mayne
Susan McReynolds
Christine Ross
Badette Tribbey
Steve Winterfeld

Análisis de datos

Chelsea Tuttle

Materiales promocionales

Annie Brunholzl

Marketing y publicación

Georgina Morales
Emily Spinks

Más información sobre el estado de Internet en materia de seguridad

Lea números anteriores del aclamado informe sobre el estado de Internet en materia de seguridad de Akamai y entérese de cuándo se publican los siguientes números. **akamai.com/soti**

Más información acerca de la investigación de Akamai sobre amenazas

Conozca los últimos análisis de inteligencia frente a amenazas, informes de seguridad e investigación sobre ciberseguridad. **akamai.com/security-research**

Acceda a los datos de este informe

Vea versiones de alta calidad de los gráficos a los que se hace referencia en este informe. Puede usar estas imágenes y hacer referencia a ellas libremente, siempre que se cite debidamente a Akamai como fuente y que se conserve el logotipo de Akamai. **akamai.com/sotidata**

Más información sobre las soluciones de Akamai

Para obtener más información sobre las soluciones de Akamai para detectar los scrapers web y protegerse de ellos, visite nuestra **Content Protector page**.

Akamai protege la experiencia de sus clientes, su personal, sus sistemas y sus datos, ayudándole a integrar la seguridad en todo lo que crea, dondequiera que lo cree o distribuya. La visibilidad de las amenazas globales que ofrece nuestra plataforma nos permite adaptar y desarrollar su estrategia de seguridad para integrar el enfoque Zero Trust, detener el ransomware, proteger las aplicaciones y las API o combatir los ataques DDoS, y le proporciona la confianza necesaria para innovar, crecer y transformar todo su entorno. Para obtener más información acerca de las soluciones de cloud computing, seguridad y distribución de contenido de Akamai, visite akamai.com y akamai.com/blog, o siga a Akamai Technologies en X, antes conocido como Twitter, y LinkedIn.
Publicado el 24 de junio.