Cómo funciona la solución de gestión de bots de Cloudflare

Cloudflare utiliza enfoques automatizados y basados en datos para gestionar los bots. El uso del aprendizaje automático, huellas digitales y otras heurísticas en un grupo seleccionado de datos de tráfico en 27 millones de dominios en nuestra red, así como el análisis de comportamiento del tráfico por zona, permite a Cloudflare puntuar de manera confiable cada solicitud en función de la probabilidad que tenga de proceder de un bot.

Nuestra solución está completamente integrada con el conjunto de soluciones de seguridad de Cloudflare, incluido el WAF y la protección DDoS, así como nuestra CDN, para que los clientes no tengan que sacrificar el rendimiento a cambio de la seguridad.

+55 (11) 3230 4523 | enterprise@cloudflare.com | www.cloudflare.com

Casos de uso

La gestión de bots de Cloudflare evita las interrupciones causadas por bots malos, lo que permite que los activos de Internet estén disponibles, protegidos y funcionando en todo momento, sin impedir el acceso de los bots buenos, como los rastreadores web de Google.

La gestión de bots de Cloudflare está diseñada para funcionar en los siguientes casos de uso:

Relleno de credenciales

Apropiación de la cuenta del usuario mediante la aplicación automática de credenciales de cuenta previamente robadas.

Acumulación de inventario

Comprar productos de forma fraudulenta para privar a clientes legítimos o revenderlos a un precio más alto.

Apropiación de contenido

Extracción y robo de información de un sitio web.

Relleno de tarjetas de crédito

Intentos de validar tarjetas de crédito robadas para luego realizar compras fraudulentas.

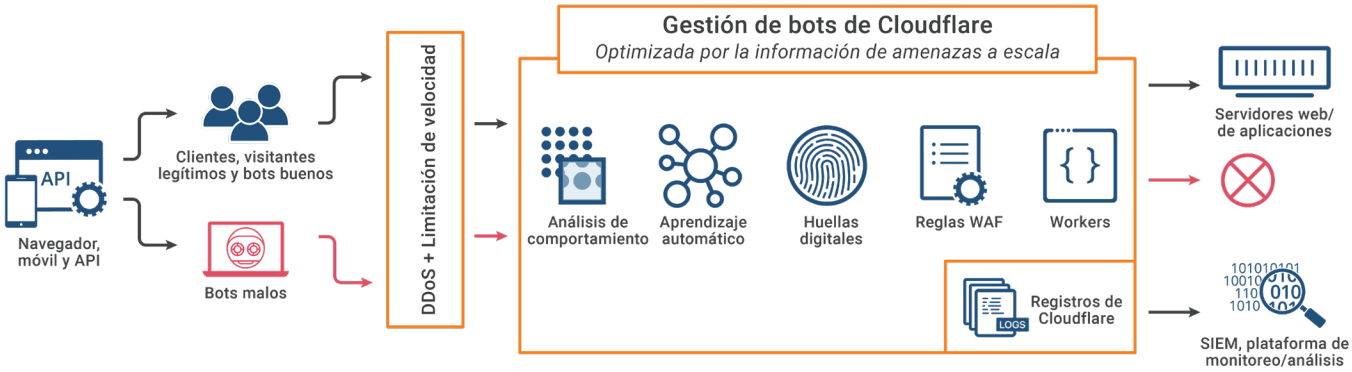
Spam de contenido

Añade contenido malicioso a propiedades web como foros y formularios de registro.

Aplicación DDoS

Ralentizar los sitios, desaprovechar ancho de banda y recursos informáticos.

Cómo funciona la gestión de bots de Cloudflare



La solución de gestión de bots de Cloudflare puede detectar con precisión el tráfico automatizado. Lo hace aprovechando la información de amenazas del tráfico en 27 millones de dominios en nuestra red. Usamos el aprendizaje automático y las huellas digitales en un grupo seleccionado de datos de ese tráfico para puntuar de manera confiable y rápida cada solicitud en función de la probabilidad que tenga de proceder de un bot. Nuestra solución también usa análisis de comportamiento para detectar anomalías en el tráfico específico del sitio, calificando cada solicitud en función de su diferencia con el

comportamiento normal. Esta puntuación se envía al motor de reglas del WAF de Cloudflare y a Cloudflare Workers, nuestro entorno de ejecución sin servidor para aplicaciones.
Según la puntuación y las reglas configuradas, el WAF puede bloquear, desafiar o registrar solicitudes sospechosas. Cloudflare Workers, por otro lado, ofrece a los clientes más flexibilidad sobre qué hacer con las solicitudes en función de la puntuación, por ejemplo, inyectar contenido nuevo o cambiar el contenido existente de la página HTML, proporcionar datos incorrectos a los bots, bloquear ciertas solicitudes, entre otras acciones.

Puntuación:

La puntuación de bots es la clave de la solución de gestión de bots de Cloudflare. Es sencilla, flexible, configurable y proporciona telemetría sobre bots por cada solicitud. Además, esta puntuación intuitiva nos permite ampliar periódicamente nuestras capacidades de detección de bots sin que los clientes necesiten ajustar ninguna configuración.

Gracias a la información de amenazas obtenida de un amplio y diverso volumen de datos distribuidos globalmente, nuestra solución determina si el tráfico del sitio está automatizado o si es hostil a las aplicaciones web. Lo hace generando una puntuación de bot, de 1 a 99, para cada solicitud HTTP entrante que llega a la red de Cloudflare.
Esta puntuación mide de forma eficaz la probabilidad de que la solicitud provenga de una fuente automatizada.

- Cuanto más alta sea la puntuación, más probabilidades hay de que la solicitud proceda de un humano que usa un navegador web estándar.

- Cuanto menor sea la puntuación, mayor será la probabilidad de que la solicitud haya sido iniciada por un script, un servicio de API, apropiadores de contenido/arañas web u otros agentes automatizados.

Puntuar cada solicitud de cada cliente tiene las siguientes ventajas:

- **Facilidad de integración:** incluso antes de habilitar la gestión de bots, podemos saber qué tan bien funcionará para el cliente específico, incluso facilitar tendencias históricas sobre la actividad de los bots.

- **Bucle de retroalimentación:** puntuar cada solicitud en la red tiene un gran valor para la mejora continua de nuestros mecanismos de detección.

- **Garantía de escalabilidad:** si podemos calcular una puntuación para cada solicitud y cliente, significa que cada propiedad de Internet en la red de Cloudflare es un cliente potencial de nuestra solución de gestión de bots.

- **Información global de los bots:** Cloudflare protege más de 27 millones de sitios web, lo que nos permite comprender y reaccionar ante los cambios tectónicos que se producen en la seguridad y la información sobre amenazas a lo largo del tiempo.

Mecanismos de detección

La plataforma de gestión de bots de Cloudflare utiliza varios mecanismos de detección, cada uno de los cuales genera una puntuación. Aunque la mayoría de los mecanismos de detección se aplican en cada solicitud, algunos están habilitados por cliente para adaptarse mejor a sus necesidades.

Análisis de comportamiento

Nuestro algoritmo de análisis de comportamiento identifica bots buscando características en todas las solicitudes. Supervisa las solicitudes siguientes e identifica bots en función de un comportamiento inusual.

El análisis de comportamiento es un enfoque de aprendizaje automático no supervisado con los siguientes beneficios:

- **Se adapta a las necesidades específicas de los clientes:** el análisis del comportamiento se habilita automáticamente para todos los clientes de nuestra solución de gestión de bots, y calcula y analiza el comportamiento normal de los visitantes durante un periodo de tiempo prolongado.

- **Detecta nuevos bots:** puede detectar nuevos bots y anomalías de comportamiento aparentemente normal en el sitio web de cada cliente.

Aprendizaje automático

Nuestros modelos de aprendizaje automático que dependen de los datos son lo que adoptan la mayoría de las decisiones sobre la puntuación final. Cloudflare da soporte a más de 27 millones de propiedades de Internet, lo que facilita una fuente abundante de diversos datos. Usamos el aprendizaje automático en cada solicitud que recibimos mediante un mecanismo de detección de aprendizaje automático llamado CatBoost, que es una biblioteca de código abierto sumamente eficaz para la potenciación del gradiente en los árboles de decisión. CatBoost ofrece las siguientes funciones clave para la gestión de bots de Cloudflare:

- **Características categóricas:** nos permiten entrenar sobre características de cardinalidad incluso muy altas.

- **Precisión superior:** nos permite reducir el sobreajuste mediante el uso de un nuevo esquema que potencia el gradiente.

- **Velocidad de inferencia:** en nuestro caso, se necesitan menos de 50 microsegundos para aplicar cualquiera de nuestros modelos, lo que garantiza que el procesamiento de solicitudes sea ultrarápido.

- **Admite C y Rust API:** la mayor parte de nuestra lógica de negocio en el perímetro se escribe utilizando LUA, más específicamente LuaJIT, por lo que tener una interfaz de FFI compatible para poder aplicar modelos es fantástico.

Hay varios modelos de CatBoost que se ejecutan en el perímetro de Cloudflare en modo virtual paralelo en *cada solicitud y en cada equipo*. Uno de los modelos se ejecuta en modo activo, lo que influye en la puntuación final que va al WAF y Workers de Cloudflare. Todos los resultados y las funciones de detección del aprendizaje automático se registran y graban en ClickHouse para llevar a cabo nuevos análisis, mejoras en el modelo, analíticas y registros de cara al cliente. Aplicamos características categóricas y numéricas en nuestros modelos, extraídas de los atributos de solicitud y características entre solicitudes creadas utilizando esos atributos, calculadas y entregadas por nuestra *plataforma de características entre solicitudes*.

Podemos implementar nuevos modelos de aprendizaje automático en cuestión de segundos utilizando una base de datos de configuración Quicksilver, muy fiable y eficaz. El mismo mecanismo se puede utilizar para decidir qué modelo de aprendizaje automático debe ejecutarse en modo activo para un cliente específico.

Motor heurístico

Nuestro motor heurístico también se aplica a cada solicitud. Tenemos distintos tipos de heurísticas y cientos de reglas específicas basadas en ciertos atributos de la solicitud, algunos de los cuales son muy difíciles de falsificar. Cuando una de las solicitudes coincide con cualquiera de las heurísticas, asignamos la puntuación más baja posible de 1.

El motor heurístico tiene las siguientes propiedades:

- **Velocidad:** si la inferencia del modelo de aprendizaje automático es inferior a 50 microsegundos por modelo, se pueden aplicar cientos de heurísticas en ¡poco menos de 20 microsegundos!
- **Implementación:** el motor heurístico nos permite añadir una nueva heurística en cuestión de segundos usando Quicksilver, y se aplicará en cada solicitud.
- **Amplia cobertura:** el uso de un conjunto de heurística simple nos permite clasificar el ~15 % del tráfico global y el ~30 % del tráfico de los clientes de nuestra solución como bots. No está tan mal para algunas expresiones condicionales "If" ¿verdad?
- **Menos falsos positivos:** debido a que somos muy conservadores con la heurística que agregamos, este mecanismo de detección tiene la tasa de falsos positivos más baja entre todos los mecanismos de detección.
- **Etiquetas para aprendizaje automático:** utilizamos solicitudes clasificadas con heurística para entrenar nuestros modelos de aprendizaje automático, que luego pueden generalizar el comportamiento aprendido de la heurística y mejorar la precisión de detección.

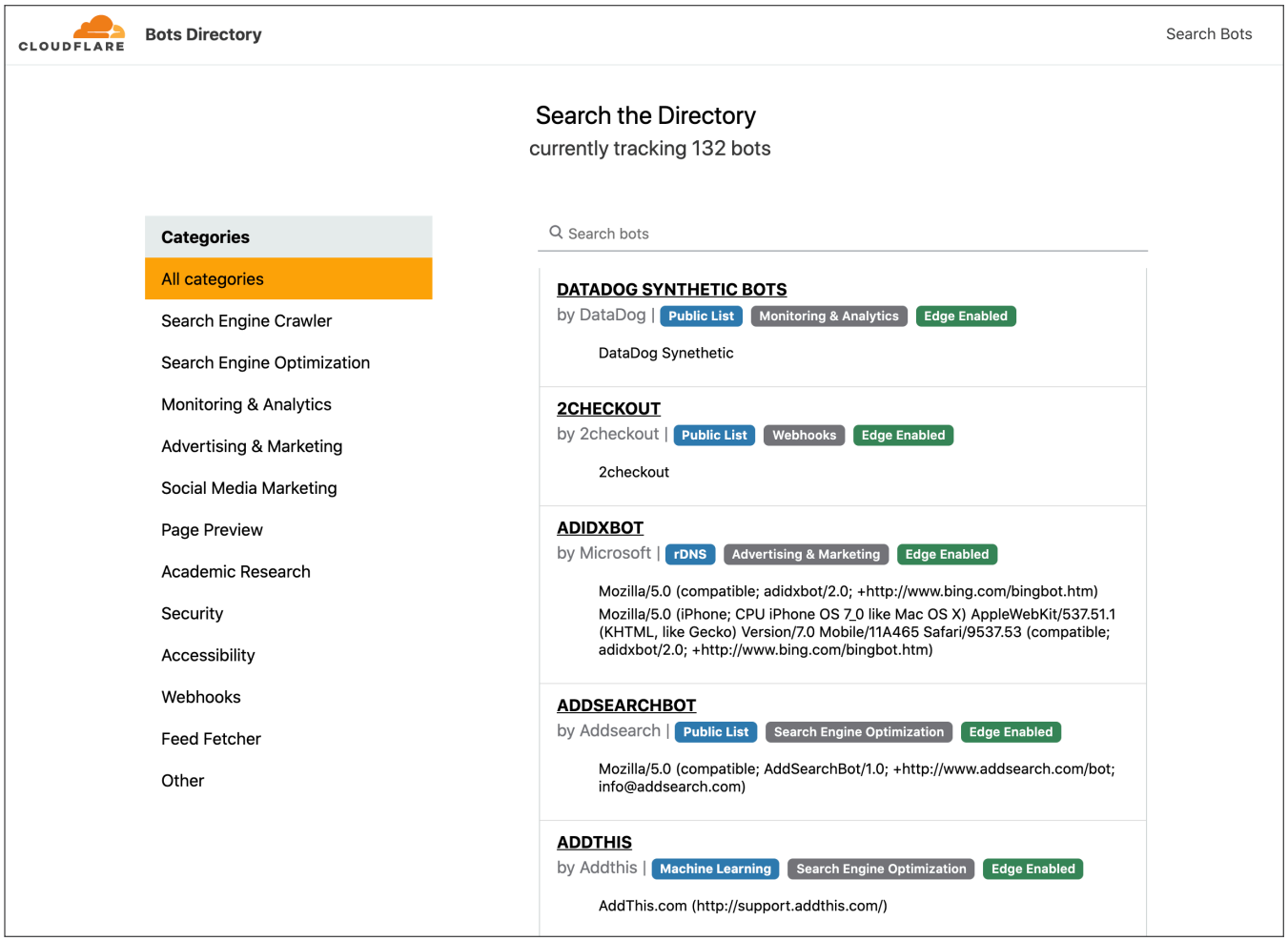
Nuestro motor heurístico tiene dos características adicionales:

i. Bot verificado

Los bots verificados son bots "buenos" conocidos, tales como Googlebot, Bingbot y LinkedInbot, que están en nuestra lista de permitidos verificada. El motor heurístico de gestión de bots de Cloudflare identifica cientos de bots "buenos" únicos que pertenecen a diferentes empresas y garantiza el acceso de estos bots útiles. Para brindar flexibilidad al usuario, todas las solicitudes de bots "buenos" por defecto obtienen una puntuación de 1. El campo Firewall permite a los clientes decidir si desean permitir el acceso de los bots "buenos" o restringirlo a ciertas partes del sitio web.

El indicador de bot verificado tiene las siguientes propiedades:

- **Un enfoque basado en validación:** admitimos múltiples mecanismos de validación, cada uno de ellos nos permite confirmar de manera confiable identidades de bots "buenos" agrupando un conjunto de direcciones IP.
- **Un validador de DNS inverso:** realiza una comprobación de DNS inverso para determinar si la dirección IP de un bot coincide o no con su supuesto nombre de servidor.
- **Un validador de bloque ASN:** similar a la verificación rDNS, pero se realiza en un bloque ASN.
- **Un validador de descarga:** recopila direcciones IP de bots buenos de archivos de texto o páginas html alojadas en los sitios de los propietarios de bots.
- **Un validador de aprendizaje automático:** utiliza un algoritmo de aprendizaje no supervisado, agrupando direcciones IP de bots buenos que no son posibles de validar por otros medios.
- **Directorio de bots:** una base de datos conectada a una interfaz de usuario que almacena y administra los bots que pasan por la red de Cloudflare.



Ejemplo de interfaz de usuario del directorio de bots

ii. Huellas digitales en Javascript

La creación de todos los gráficos del navegador web, como canvas, depende de varias capas, como el hardware (unidad de procesado gráfico) y el software (controladores, procesamiento del sistema operativo). Este resultado excepcional permite una diferenciación precisa entre diferentes tipos de navegadores/dispositivos. Además, se puede lograr sin sacrificar la privacidad de los visitantes del sitio web. No es una supercookie y no se puede utilizar para rastrear e identificar usuarios individuales. Se utiliza solo para confirmar que el agente de usuario de una solicitud coincida con otros datos de telemetría recopilados a través de la API de canvas.

Este mecanismo de detección se implementa como un sistema de desafío-respuesta, en el que el desafío se inyecta en la página web en el perímetro de Cloudflare. El desafío se representa a continuación en segundo plano utilizando las instrucciones gráficas proporcionadas y el resultado se envía de vuelta a Cloudflare para su validación y para realizar otras acciones, como generar la puntuación. La huella digital en JS asegura resultados confiables que son impermeables a los ataques de repetición, todo ello sin sacrificar la privacidad del usuario.

Una vez que tengamos la puntuación combinada de todos estos motores, estará disponible en: reglas de firewall y Cloudflare Workers.

Integración WAF

El panel intuitivo del WAF de Cloudflare permite a los usuarios crear reglas eficaces con solo presionar un botón y también permite la integración con Terraform. Cada solicitud al WAF se inspecciona conforme al motor de reglas. Las solicitudes sospechosas se pueden bloquear, desafiar o registrar según las necesidades del usuario, mientras que las solicitudes legítimas se enrutan al destino, según la puntuación presentada por el módulo de gestión de bots y el umbral configurado.

Las reglas de firewall ofrecen las siguientes acciones de mitigación de bots:

- **Registrar:** registra las solicitudes que coinciden en los registros de Cloudflare.
- **Omitir:** permite a los clientes desactivar dinámicamente las funciones de seguridad de Cloudflare para una solicitud.
- **Permitir:** las solicitudes coincidentes están exentas de desafíos y bloquean las acciones activadas por otro contenido de reglas de firewall.
- **Desafiar (CAPTCHA):** útil para garantizar que el visitante que accede al sitio sea humano y no automatizado.
- **Resolver desafío JS:** útil para garantizar que los bots malos no puedan acceder al recurso solicitado. Los navegadores, sin embargo, son libres de superar el desafío automáticamente.
- **Bloquear**: a las solicitudes coincidentes se les niega el acceso al sitio.

Integración Workers

Cloudflare Workers ofrece a los clientes más flexibilidad sobre qué hacer con las solicitudes en función de la puntuación. La implementación más sencilla con Workers sería enviar la puntuación de bot como encabezado de solicitud para las solicitudes al origen. Otras implementaciones basadas en Workers para la gestión de bots incluyen, entre otras, las siguientes:

- Capturar la puntuación con una solicitud.
- Entregar contenido alternativo inyectando contenido nuevo o cambiando el contenido existente de la página HTML.
- Redirigir la solicitud (a otra página, aplicación, honeypot, cómputo, etc.).
- Ofrecer contenido diferente según la puntuación (no publiques anuncios con bots con una puntuación baja).
- Exigir autorización adicional (por ejemplo, autorización de token para solicitudes con baja puntuación).
- Detener ciertas solicitudes
- Comprobar combinaciones basadas en la puntuación del bot y la cookie de autorización (mostrar qué bots aprobaron el desafío CAPTCHA, o qué humanos obtuvieron puntuaciones bajas).

Por ejemplo, utilizando este pequeño fragmento de código, podemos devolver la puntuación al servidor de origen para un análisis o mitigación más avanzada en tiempo real:

```
addEventListener('fetch', event => {
  event.respondWith(handleRequest(event.request))
})

async function handleRequest(request) {
  request = new Request(request);

  request.headers.set("Cf-Bot-Score", request.cf.bot _ management.score)

  return fetch(request);
}
```

Informes y análisis:

Cloudflare mantiene registros detallados de todos los datos de solicitud. El resumen general básico de las métricas relacionadas con los bots está disponible a través de la integración de Cloudflare Firewall Analytics, e incluye el número de amenazas detectadas, el número de desafíos CAPTCHA generados y resueltos, y el origen de cada ataque (ASN, país, IP, agente de usuario), lo que permite a los clientes calcular tasas de falsos positivos.

Nuestra herramienta Firewall Analytics, que utiliza ClickHouse y la API de GraphQL, permite a los clientes identificar e investigar rápidamente las amenazas de seguridad mediante una interfaz intuitiva. Además, para el análisis, proporcionamos registros detallados de toda la actividad relacionada con los bots a través de la API de Logpull o LogPush, que permiten guardar fácilmente tus registros en tu almacenamiento en la nube. Cloudflare también ofrece análisis de la tasa de resolución de CAPTCHA y se pueden crear paneles de control más avanzados utilizando la función de Log Share. Los clientes pueden así integrar la información de amenazas de Cloudflare con sus propias herramientas de gestión de información y eventos de seguridad (SIEM).

Para ayudar en este proceso, Cloudflare ofrece integraciones "establecidas" para una variedad de proveedores (https://developers.cloudflare.com/logs/analytics-integrations/), incluidos Datadog, Elastic, Google Cloud, Looker, Splunk y Sumo Logic. Los clientes con otras herramientas SIEM pueden confiar en la herramienta flexible Log Share para crear sus propios paneles personalizados.

Además, brindamos a los clientes acceso a la puntuación de bot en sus encabezados, que se puede rastrear en herramientas de análisis, lo que permite la elaboración de informes detallados y correlación de las puntuaciones con otros factores.

Glosario

Relleno de credenciales

En los ataques de relleno de credenciales, también conocidos como ataques de apropiación fraudulenta de cuenta, los bots malos utilizan credenciales robadas de fallas de seguridad anteriores para apropiarse de cuentas y robar datos confidenciales. Esta práctica funciona porque los usuarios a menudo reutilizan credenciales en numerosos sitios: correo electrónico personal, cuentas bancarias, cuentas de tarjetas de crédito, sitios de comercio electrónico, etc. Si un atacante adquiere las credenciales de una de estas cuentas, las probará en varios sitios web. Si el usuario ha reutilizado las mismas credenciales en cualquiera de estos otros sitios, el atacante tendrá acceso a todas esas cuentas.

Acumulación de inventario

Los bots de acumulación de inventario están diseñados para interrumpir los sitios de comercio electrónico mediante la manipulación de la disponibilidad del inventario. Muchos bots compran productos cuya oferta es limitada con el objetivo de revenderlos a un precio más alto en otros sitios. Esta práctica afecta la reputación de la marca comercial, ya que los clientes no pueden comprar el producto en su sitio y se ven obligados a pagar un precio más alto en otros. Algunos bots pueden simplemente rellenar y abandonar las cestas de compras para impedir que los clientes legítimos realicen una compra. También pueden agregar repetidamente artículos muy demandados a los carritos de compras en línea sin siquiera completar una compra. El objetivo es agotar artificialmente el inventario del minorista, con el consiguiente impacto negativo tanto en las ventas como en la experiencia del cliente.

Apropiación de contenido

La apropiación de contenido o los bots de extracción de datos de un sitio web rastrean continuamente los sitios web para robar contenido, tales como datos de precios, descripciones de productos, texto, imágenes, código HTML, código CSS, etc. Estos bots a menudo se utilizan para reutilizar contenido con fines malintencionados, p.ej. crear una copia ilícita de un sitio web para defraudar a los usuarios. Los bots de extracción más sofisticados también pueden usar JavaScript para cumplimentar todos los formularios de un sitio web y descargar cualquier contenido restringido.

Relleno de tarjetas de crédito

En ataques de relleno de tarjetas de crédito, los bots automatizados intentan validar las tarjetas de crédito robadas con la esperanza de realizar compras fraudulentas, transferir dinero o retirar sumas de esas tarjetas. Antes de realizar transacciones más grandes o vender las tarjetas validadas en la dark web (Internet oscuro), los piratas informáticos utilizan tarjetas de crédito robadas para realizar compras de valor pequeño en sitios web menos seguros. El comercio y el titular de la tarjeta a menudo no las detectan hasta que es demasiado tarde. Las compras se realizan o se transfiere el dinero. El pago de reembolsos a los clientes y la pérdida de mercancía en transacciones fraudulentas pueden tener graves impactos financieros para los comerciantes. Además, una tasa de fraude elevada podría llevar a una compañía de tarjetas de crédito a penalizar al comercio o suspender su colaboración con ellos.

Spam de contenido

Los bots de spam de contenido rellenan los formularios en redes sociales con un propósito hostil. Pueden publicar comentarios y reseñas falsas, contenido inapropiado o incluso publicitar ofertas demasiado buenas para ser ciertas con el fin de atraer a los usuarios a hacer clic en enlaces a sitios web maliciosos que pueden favorecer ataques de phishing o de intermediario. Sus intenciones pueden ser expulsar a los clientes enviándolos a otros sitios, propagar malware o sabotear la reputación de un negocio. Las publicaciones en redes sociales y las reseñas de productos pueden influir de forma importante en los consumidores y el contenido falso puede sabotear la reputación de un negocio. Los bots de extracción de contenido más sofisticados pueden usar JavaScript para, por ejemplo, rellenar cada formulario en un sitio web, descargar cualquier contenido restringido y reutilizarlo sin darte ningún reconocimiento.

Ataques DDoS a aplicaciones

Si observas todos los ataques que hemos mencionado hasta ahora (excepto la extracción de contenido con una buena memoria caché), requieren solicitudes de servidor de origen, que consumen recursos. Demasiadas solicitudes pueden interrumpir una aplicación al sobrecargar el servidor de origen, crear latencia e incluso, a veces, desconectar el servidor. Los atacantes pueden crear de manera intencionada o involuntaria este tipo de interrupción utilizando varios clientes bot para saturar el servidor de solicitudes, interrumpiendo así el tráfico de los usuarios legítimos.

Los bots malos pueden impactar en tus ingresos, arruinar la reputación de tu marca, ralentizar la experiencia del usuario final, aumentar los costos de tu infraestructura e interrumpir por completo tu servicio. La solución de gestión de bots de Cloudflare puede ayudar a identificar y bloquear estos bots malos, manteniéndote a salvo.

Conclusión

Cloudflare utiliza enfoques automatizados y basados en datos para gestionar los bots. El uso del aprendizaje automático, huellas digitales y otras heurísticas en un grupo seleccionado de datos de tráfico en 27 millones de dominios en nuestra red, así como el análisis de comportamiento del tráfico por zona, permite a Cloudflare puntuar de manera confiable cada solicitud en función de la probabilidad que tenga de proceder de un bot. Nuestra solución está completamente integrada con el conjunto de soluciones de seguridad de Cloudflare, incluido el WAF y la protección DDoS, así como nuestra CDN, para que los clientes no tengan que sacrificar el rendimiento a cambio de la seguridad. La implementación y la administración de la solución de gestión de bots de Cloudflare basada en la nube es fácil.

Visita **www.cloudflare.com/products/bot-management**
para obtener más información o comunícate con tu gerente de cuenta de Cloudflare.

CLOUDFLARE®

+55 (11) 3230 4523 | enterprise@cloudflare.com | www.cloudflare.com

REV: 200820