Documento técnico de AWS

Prácticas recomendadas de AWS para la resiliencia DDoS

Prácticas recomendadas de AWS para la resiliencia DDoS: Documento técnico de AWS

Table of Contents

Prácticas recomendadas de AWS para la resiliencia frente a ataques DDoS

Fecha de publicación: 21 de septiembre de 2021 (Revisiones del documento)

Resumen

Es importante proteger su empresa del impacto de los ataques de denegación de servicio distribuido (DDoS, por sus siglas en inglés), así como de otros ciberataques. Mantener la confianza del cliente en su servicio y conservar la disponibilidad y la capacidad de respuesta de su aplicación es su máxima prioridad. Además de evitar gastos directos innecesarios en el caso de que deba reducir su infraestructura horizontalmente en respuesta a un ataque. Amazon Web Services (AWS) se compromete a proporcionarle las herramientas, las prácticas recomendadas y los servicios para defenderse de los agentes malintencionados en Internet. Usar los servicios correctos de AWS ayuda a garantizar una elevada disponibilidad, seguridad y resiliencia.

En este documento técnico, AWS proporciona orientación prescriptiva sobre la DDoS para mejorar la resiliencia de las aplicaciones que se ejecutan en AWS. Aquí se incluye una arquitectura de referencia resistente a DDoS que se puede utilizar como guía para ayudar a proteger la disponibilidad de las aplicaciones. Este documento técnico también describe diferentes tipos de ataques, como los ataques a la capa de infraestructura y los ataques a la capa de aplicación. AWS explica qué prácticas recomendadas son las más eficaces a la hora de gestionar cada tipo de ataque. Además, se describen los servicios y las características que se ajustan a una estrategia de mitigación de DDoS y se explica cómo se puede usar cada servicio para ayudar a proteger sus aplicaciones.

Este documento está dirigido a los responsables de la toma de decisiones de TI y a los ingenieros de seguridad familiarizados con los conceptos básicos de redes, seguridad y AWS. Cada sección tiene enlaces a la documentación de AWS, donde se proporcionan más detalles sobre las prácticas recomendadas o las capacidades.

Introducción: ataques de denegación de servicio

Un ataque de denegación de servicio (DoS, por sus siglas en inglés) es un intento deliberado de hacer que un sitio web o una aplicación no estén disponibles para los usuarios, por ejemplo, inundándolo de tráfico de red. Los atacantes utilizan una variedad de técnicas que consumen grandes cantidades de ancho de banda de red o bloquean otros recursos del sistema, lo que interrumpe el acceso de los usuarios legítimos. En su forma más simple, un único atacante utiliza una sola fuente para llevar a cabo un ataque DoS contra un objetivo, tal como se muestra en la siguiente imagen.

Tabla 1: Diagrama del ataque DoS

En un ataque DDoS, un atacante utiliza varios orígenes para orquestar un ataque contra un objetivo. Estos orígenes pueden incluir grupos distribuidos de equipos, enrutadores, dispositivos de IoT y otros puntos de conexión infectados con malware. En el siguiente diagrama se muestra una red de hosts comprometidos que participa en el ataque, lo que genera una inundación de paquetes o solicitudes para abrumar al objetivo.

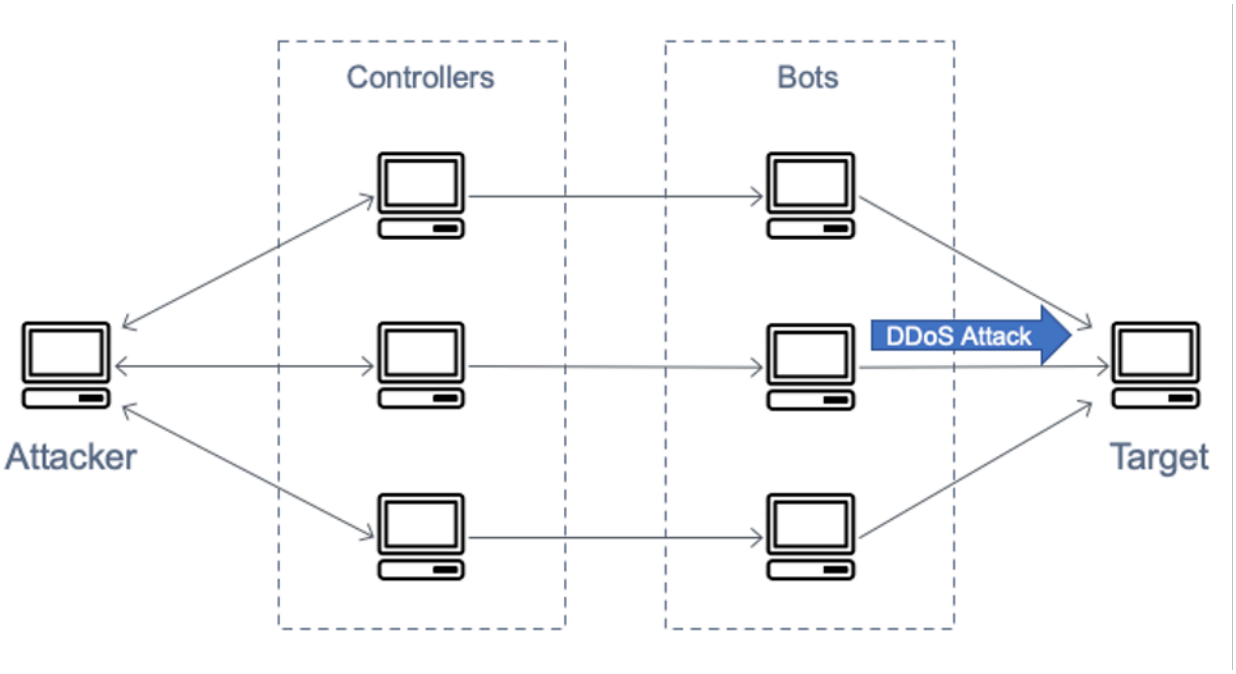

Diagrama del ataque DDoS

Hay siete capas en el modelo de interconexión de sistemas abiertos (OSI, por sus siglas en inglés) que se describen en la tabla Modelo de interconexión de sistemas abiertos (OSI). Los ataques

DDoS son más comunes en las capas tres, cuatro, seis y siete. Los ataques de las capas tres y cuatro corresponden a las capas de red y transporte del modelo OSI. En este documento, AWS hace referencia a ellos colectivamente como ataques a la capa de infraestructura. Los ataques de las capas seis y siete corresponden a las capas de presentación y aplicación del modelo OSI. AWS designa a estos ataques conjuntamente como ataques a la capa de aplicación. En las siguientes secciones se analizan ejemplos de estos tipos de ataques.

Modelo de interconexión de sistemas abiertos (OSI)

Número	Capa	Unidad	Descripción	Ejemplos de vector
7	Aplicación	Datos	Procesamiento de red para la aplicación	Inundaciones de HTTP, inundaciones de consultas de DNS
6	Presentación	Datos	Representación de datos y cifrado	Abuso de TLS
5	Sesión	Datos	Comunicación entre hosts	N/A
4	Transporte	Segmentos	Conexiones de extremo a extremo y fiabilidad	Inundaciones de SYN
3	Red	Paquetes	Determinación de la ruta y direccionamiento lógico	Ataques de reflexión UDP
2	Enlace de datos	Fotogramas	Direccionamiento físico	N/A

Número	Capa	Unidad	Descripción	Ejemplos de vector
1	Física	Bits	Transmisión multimedia, de señal y binaria	N/A

Temas

Ataques a la capa de infraestructura

Los ataques más comunes a la capa de infraestructura son los ataques DDoS, los ataques de reflexión del protocolo de datagramas de usuario (UDP, por sus siglas en inglés) y las inundaciones de sincronización (SYN). Un atacante puede usar cualquiera de estos métodos para generar grandes volúmenes de tráfico que pueden inundar la capacidad de una red o inmovilizar recursos en sistemas tales como servidores, firewalls, sistemas de prevención de intrusiones (IPS, por sus siglas en inglés) o equilibradores de carga. Si bien estos ataques se pueden identificar fácilmente, para mitigarlos de manera efectiva se necesita una red o sistemas que escalen verticalmente la capacidad más rápidamente que la inundación de tráfico entrante. Esta capacidad adicional es necesaria para filtrar o absorber el tráfico de ataque, liberando el sistema y la aplicación para responder al tráfico legítimo de los clientes.

Temas

Ataques de reflexión UDP

Los ataques de reflexión del protocolo de datagramas de usuario (UDP) aprovechan el hecho de que UDP es un protocolo sin estado. Los atacantes pueden crear un paquete de solicitud UDP válido que incluya la dirección IP del objetivo del ataque como la dirección IP de origen del UDP. El atacante ha falsificado (suplantado) la IP de origen del paquete de solicitud UDP. El paquete UDP contiene la IP de origen suplantada y el atacante lo envía a un servidor intermedio. Se engaña al servidor para

que envíe los paquetes de respuesta UDP a la IP de la víctima objetivo en lugar de devolverlos a la dirección IP del atacante. El servidor intermedio se utiliza porque genera una respuesta que es varias veces mayor que el paquete de solicitud, lo que amplifica de manera efectiva la cantidad de tráfico que envía el atacante a la dirección IP de destino.

El factor de amplificación es la relación entre el tamaño de la respuesta y el tamaño de la solicitud y varía según el protocolo que utilice el atacante: DNS, NTP, SSDP, CLDAP, Memcached, CharGen o QOTD. Por ejemplo, el factor de amplificación para el DNS puede ser de 28 a 54 veces el número original de bytes. Por lo tanto, si un atacante envía una carga de solicitud de 64 bytes a un servidor DNS, puede generar más de 3400 bytes de tráfico no deseado a un objetivo de ataque. Los ataques de reflexión UDP son responsables de un mayor volumen de tráfico en comparación con otros ataques. En la ilustración Ataque de reflexión UDP se muestra la táctica de reflexión y el efecto de amplificación.

Ataque de reflexión UDP

Ataques de inundación SYN

Cuando un usuario se conecta a un servicio de protocolo de control de transmisión (TCP, por sus siglas en inglés), como un servidor web, su cliente envía un paquete de sincronización SYN. El servidor devuelve un paquete SYN-ACK como acuse de recibo y, por último, el cliente responde con un paquete de acuse de recibo (ACK), que completa el apretón de manos de tres vías esperado. La siguiente imagen ilustra el típico apretón de manos.

Apretón de manos de 3 vías SYN

En un ataque de inundación SYN, un cliente malintencionado envía una gran cantidad de paquetes SYN, pero nunca envía los paquetes ACK finales para completar los apretones de manos. El servidor se queda esperando una respuesta a las conexiones TCP medio abiertas y, al final, se queda sin capacidad para aceptar nuevas conexiones TCP. Esto puede impedir que no se puedan conectar usuarios nuevos al servidor. El ataque intenta vincular las conexiones de servidor disponibles para que los recursos no estén disponibles para las conexiones legítimas. Si bien las inundaciones SYN pueden alcanzar hasta cientos de Gbps, el objetivo del ataque no es aumentar el volumen de tráfico SYN.

Ataques a la capa de aplicación

Un atacante puede apuntar hacia la propia aplicación mediante un ataque de capa 7 o de capa de aplicación. En estos ataques, similares a los ataques a la infraestructura de inundación SYN, el atacante intenta sobrecargar funciones específicas de una aplicación para que la aplicación no esté disponible o no responda a los usuarios legítimos. En ocasiones, esto se puede lograr con volúmenes de solicitudes muy bajos que generan solo un volumen pequeño de tráfico de red. Esto puede hacer que el ataque sea difícil de detectar y no se pueda mitigar. Los ejemplos de ataques a la capa de aplicación incluyen inundaciones de HTTP, ataques de eliminación de caché e inundaciones XML-RPC de WordPress.

En un ataque de inundación de HTTP, un atacante envía solicitudes HTTP que parecen provenir de un usuario válido de la aplicación web. Algunas inundaciones de HTTP se dirigen a un recurso específico, mientras que las inundaciones de HTTP más complejas intentan emular la interacción humana con la aplicación. Esto puede aumentar la dificultad de usar técnicas de mitigación comunes, como la limitación de la velocidad de solicitudes.

Los ataques de eliminación de caché son un tipo de inundación de HTTP que utiliza variaciones en la cadena de consulta para eludir el almacenamiento en caché de la red de entrega de contenido (CDN, por sus siglas en inglés). En lugar de poder devolver los resultados almacenados en caché, la CDN debe contactar con el servidor de origen para cada solicitud de página, y estas recuperaciones de origen provocan una tensión adicional en el servidor web de la aplicación.

Con un ataque de inundación XML-RPC de WordPress, también conocido como inundación de pingback de WordPress, un atacante ataca un sitio web alojado en el software de gestión de contenido de WordPress. El atacante usa incorrectamente la función de API XML-RPC para generar una inundación de solicitudes HTTP. La característica de pingback permite que un sitio web alojado en WordPress (sitio A) notifique a un sitio diferente de WordPress (sitio B) a través de un enlace que el sitio A ha creado al sitio B. El sitio B intenta buscar el sitio A para verificar la existencia del enlace. En una inundación de pingback, el atacante hace un mal uso de esta capacidad para hacer que el sitio B ataque al sitio A. Este tipo de ataque tiene una firma clara: WordPress suele estar presente en el agente de usuario del encabezado de la solicitud HTTP.

Existen otras formas de tráfico malintencionado que pueden afectar a la disponibilidad de una aplicación. Los bots Scraper automatizan los intentos de acceder a una aplicación web para robar contenido o registrar información competitiva, como los precios. Los ataques de fuerza bruta y la reutilización de credenciales robadas son esfuerzos programados para obtener acceso no autorizado a áreas seguras de una aplicación. No se trata estrictamente de ataques DDoS; pero su naturaleza

automatizada puede parecerse a un ataque DDoS y se pueden mitigar implementando algunas de las mismas prácticas recomendadas que se explican en este documento.

Los ataques a la capa de aplicación también pueden dirigirse a los servicios del sistema de nombres de dominio (DNS, por sus siglas en inglés). El más común de estos ataques es una inundación de consultas de DNS en la que un atacante utiliza muchas consultas de DNS bien formadas para agotar los recursos de un servidor de DNS. Estos ataques también pueden incluir un componente de eliminación de caché en el que el atacante aleatoriza la cadena de subdominio para evitar la memoria caché de DNS local de cualquier solucionador determinado. En consecuencia, el solucionador no puede aprovechar las consultas de dominio almacenadas en caché y, en su lugar, debe contactar repetidamente con el servidor de DNS autorizado, lo que amplifica el ataque.

Si una aplicación web se entrega a través de la seguridad de la capa de transporte (TLS, por sus siglas en inglés), un atacante también puede elegir atacar el proceso de negociación de TLS. La TLS es cara desde el punto de vista computacional, por lo que un atacante, al generar una carga de trabajo adicional en el servidor para procesar datos ilegibles (o ininteligibles [texto cifrado]) como un apretón de manos legítimo, puede reducir la disponibilidad del servidor. Otra variante posible de este ataque es que un atacante complete el protocolo de apretón de manos de TLS pero renegocie permanentemente el método de cifrado. Alternativamente, un atacante puede intentar agotar los recursos del servidor abriendo y cerrando muchas sesiones de TLS.

Técnicas de mitigación

Los servicios de AWS incluyen algunas formas de mitigación de DDoS de forma automática. La resiliencia de DDoS se puede mejorar aún más mediante el uso de una arquitectura de AWS con servicios específicos, que se tratan en las siguientes secciones, y mediante la implementación de otras prácticas recomendadas para cada parte del flujo de red entre los usuarios y su aplicación.

Todos los clientes de AWS se pueden beneficiar de la protección automática de AWS Shield Standard sin cargo adicional. AWS Shield Standard protege frente a los ataques DDoS más comunes que se suelen producir en la capa de red y de transporte dirigidos contra su sitio web o sus aplicaciones. Esta protección está siempre activa, preconfigurada, estática y no proporciona informes ni análisis. Se ofrece en todos los servicios de AWS y en todas las regiones de AWS. En las regiones de AWS, los ataques DDoS se detectan y el sistema Shield Standard establece automáticamente la línea de base del tráfico, identifica anomalías y crea mitigaciones, según sea necesario. Se puede utilizar AWS Shield Standard como parte de una arquitectura resistente a DDoS para proteger tanto las aplicaciones web como las no web.

También se pueden utilizar los servicios de AWS que operan desde ubicaciones de borde, como Amazon CloudFront, Global Accelerator y Route 53 para crear una protección de disponibilidad integral contra todos los ataques conocidos a la capa de infraestructura. Estos servicios forman parte de la red global de borde de AWS y pueden mejorar la resiliencia ante DDoS de su aplicación cuando atienden cualquier tipo de tráfico de aplicaciones desde ubicaciones de borde distribuidas en todo el mundo. Puede ejecutar su aplicación en cualquier región de AWS y utilizar estos servicios para proteger la disponibilidad de su aplicación y optimizar el rendimiento de su aplicación para los usuarios finales legítimos.

Los beneficios de usar Amazon CloudFront, Global Accelerator y Amazon Route 53 son:

- Acceso a Internet y capacidad de mitigación de DDoS en toda la red global de borde de AWS. Esto es útil para mitigar ataques volumétricos mayores, que pueden llegar a alcanzar los terabits.
- Los sistemas de mitigación de DDoS de AWS Shield se integran con los servicios de borde de AWS, lo que reduce el tiempo de mitigación de minutos a subsegundos.
- Las técnicas de mitigación de inundaciones SYN sin estado representan y verifican las conexiones entrantes antes de pasarlas al servicio protegido. De este modo puede garantizar que solo lleguen a su aplicación las conexiones válidas y, al mismo tiempo, proteger a sus usuarios finales legítimos contra caídas de falsos positivos.

- Sistemas automáticos de ingeniería de tráfico que dispersan o aíslan el impacto de grandes ataques DDoS volumétricos. Todos estos servicios aíslan los ataques en la fuente antes de que lleguen a su origen, lo que significa un menor impacto en los sistemas protegidos por dichos servicios.
- La defensa de la capa de aplicación, cuando se combina con AWS WAF, no requiere cambiar la arquitectura de la aplicación actual (por ejemplo, en una región de AWS o en un centro de datos local).

No se cobra por la transferencia de datos entrante en AWS y no se paga por el tráfico de ataque DDoS mitigado por AWS Shield. El siguiente diagrama de arquitectura muestra los servicios de la red de global de borde de AWS.

Esta arquitectura incluye varios servicios de AWS que pueden ayudarlo a mejorar la resiliencia de su aplicación web contra los ataques DDoS. En la tabla Resumen de las prácticas recomendadas se incluye un resumen de estos servicios y las capacidades que pueden proporcionar. AWS ha etiquetado cada servicio con un indicador de prácticas recomendadas (BP1, BP2) para facilitar la consulta en este documento. Por ejemplo, en una sección posterior se analizan las capacidades proporcionadas por Amazon CloudFront y Global Accelerator, que incluyen el indicador de prácticas recomendadas BP1.

Tabla 2 - Resumen de las prácticas recomendadas

Borde de AWS	Región de AWS					
	Si se usa Amazon CloudFront (BP1) con AWS WAF (BP2)	Si se usa Global Accelerator (BP1)	Si se usa Amazon Route 53 (BP3)	Si se usa Elastic Load Balancing (BP6) con AWS WAF (BP2)	Si se usan grupos de seguridad y ACL de red en Amazon VPC (BP5)	Si se usa Amazon EC2 Auto Scaling (BP7)
Mitigación de ataques de capa	✓	✓	✓	✓	✓	✓

Borde de AWS	Región de AWS					
3 (por ejemplo, reflexión UDP)						
Mitigación de ataques de capa 4 (por ejemplo, inundación SYN)	✓	✓	✓	✓		
Mitigación de ataques de capa 6 (por ejemplo, TLS)	✓	✓	✓	✓		
Reducir la superficie de ataque	✓	✓	✓	✓	✓	
Escalar para absorber el tráfico de la capa de aplicación	✓	✓	✓	✓	✓	✓

Borde de AWS	Región de AWS					
Mitigación de ataques de capa 7 (capa de aplicación)	✓	✓(*)	✓	✓	✓(*)	✓(*)
Aislamiento geográfico y dispersión del exceso de tráfico y ataques DDoS mayores	✓	✓	✓			
✓ (*): si se usa con AWS WAF con el Application Load Balancer						

Otra forma de mejorar su preparación para responder y mitigar los ataques DDoS es suscribirse a AWS Shield Advanced.

Los clientes reciben una detección personalizada basada en:

- Patrones de tráfico específicos de su aplicación.
- Protección contra ataques DDoS de capa 7, incluido AWS WAF, sin coste adicional.
- Acceso a soporte especializado de forma ininterrumpida desde AWS SRT.
- Gestión centralizada de políticas de seguridad mediante AWS Firewall Manager.

- Protección de costes para protegerse de cargos de escalado resultantes de los picos de uso relacionados con DDoS.

Este servicio de mitigación de DDoS opcional ayuda a proteger las aplicaciones alojadas en cualquier región de AWS. El servicio está disponible en todo el mundo para CloudFront, Route 53 y Global Accelerator. El uso de Shield Advanced con direcciones IP elásticas le permite proteger instancias de Network Load Balancer (NLB) o de Amazon EC2.

Los beneficios de usar AWS Shield Advanced son:

- Acceso a AWS SRT para obtener ayuda para mitigar los ataques DDoS que afectan a la disponibilidad de las aplicaciones.
- La visibilidad de los ataques DDoS mediante la AWS Management Console, las API, y las métricas y alarmas de Amazon CloudWatch.
- Acceso al historial de todos los eventos DDoS de los últimos 13 meses.
- Acceso al firewall de aplicaciones web de AWS (AWS WAF), sin coste adicional, para la mitigación de los ataques DDoS en la capa de aplicación (cuando se usa con Amazon CloudFront o Application Load Balancer).
- Línea de base automática de los atributos de tráfico web, cuando se usa con AWS WAF.
- Acceso a AWS Firewall Manager, sin coste adicional, para la aplicación automatizada de políticas.
- Umbrales de detección sensibles que dirigen el tráfico al sistema de mitigación de DDoS con antelación y pueden mejorar el tiempo para mitigar los ataques contra Amazon EC2 o Network Load Balancer, cuando se usan con una dirección IP elástica.
- Protección de costes que le permite solicitar un reembolso limitado de los costes relacionados con el escalado que resultan de un ataque DDoS.
- Acuerdo de nivel de servicio mejorado que es específico de los clientes de AWS Shield Advanced.
- Participación proactiva de AWS SRT cuando se detecta un evento Shield.
- Grupos de protección que le permiten agrupar recursos, lo que proporciona una forma de autoservicio para personalizar el alcance de la detección y la mitigación de su aplicación al tratar varios recursos como una sola unidad. La agrupación de recursos mejora la precisión de la detección, minimiza los falsos positivos, facilita la protección automática de los recursos recién creados y acelera el tiempo para mitigar los ataques contra muchos recursos que incluyen una sola aplicación. Para obtener información sobre los grupos de protección, consulte Grupos de protección de Shield Advanced.

Para obtener una lista completa de las características de AWS Shield Advanced y más información sobre AWS Shield, consulte el tema Cómo funciona AWS Shield.

Temas

Prácticas recomendadas para mitigar los ataques DDoS

En las siguientes secciones, se describen con más profundidad las prácticas recomendadas para mitigar ataques DDoS. Para obtener una guía rápida y fácil de implementar sobre la creación de una capa de mitigación de DDoS para aplicaciones web estáticas o dinámicas, consulte el tema sobre cómo ayudar a proteger las aplicaciones web dinámicas contra los ataques DDoS.

Defensa de la capa de infraestructura (BP1, BP3, BP6, BP7)

En un entorno de centro de datos tradicional, los ataques DDoS en la capa de infraestructura se pueden mitigar aplicando técnicas como el aprovisionamiento excesivo de capacidad, la implementación de sistemas de mitigación de DDoS o la depuración del tráfico con la ayuda de los servicios de mitigación de DDoS. En AWS, las capacidades de mitigación de DDoS se incluyen de forma automática; pero puede optimizar la resiliencia de DDoS de su aplicación al tomar decisiones de arquitectura que aprovechen mejor esas capacidades y que, además, permitan escalar el exceso de tráfico.

Las consideraciones clave para ayudar a mitigar los ataques DDoS volumétricos son garantizar que haya suficiente capacidad y diversidad de tránsito y proteger los recursos de AWS, como las instancias de Amazon EC2, contra el tráfico de ataque.

Algunos tipos de instancias de Amazon EC2 admiten características que pueden gestionar con mayor facilidad grandes volúmenes de tráfico, por ejemplo, interfaces de ancho de banda de red de hasta 100 Gbps y redes mejoradas. Esto ayuda a evitar la congestión de la interfaz del tráfico que ha llegado a la instancia de Amazon EC2. Las instancias que admiten redes mejoradas proporcionan un mayor rendimiento de E/S, un mayor ancho de banda y una menor utilización de la CPU en comparación con las implementaciones tradicionales. Esto mejora la capacidad de la instancia para gestionar grandes volúmenes de tráfico y, en última instancia, hace que sea muy resistente a la carga de paquetes por segundo (pps).

Para permitir este alto nivel de resiliencia, AWS recomienda el uso de instancias dedicadas de Amazon EC2 o instancias de Amazon EC2 con un rendimiento de red mayor, que tengan un sufijo N y que admitan redes mejoradas con un ancho de banda de red de hasta 100 Gbps, por ejemplo, c6gn.16xlarge y c5n.18xlarge o instancias metal (como c5n.metal).

Para obtener más información sobre las instancias de Amazon EC2 que admiten interfaces de red de 100 Gigabit y redes mejoradas, consulte el tema Tipos de instancias de Amazon EC2.

El módulo necesario para las redes mejoradas y el conjunto de atributos enaSupport requerido se incluyen con Amazon Linux 2 y las versiones más recientes de la AMI de Amazon Linux. Por lo tanto, si lanza una instancia con una versión HVM de Amazon Linux en un tipo de instancia admitido, la red mejorada ya estará habilitada para su instancia. Para obtener más información, consulte el tema Probar si las redes mejoradas están habilitadas. Para obtener más información sobre cómo habilitar las redes mejoradas, consulte Redes mejoradas en Linux.

Amazon EC2 con Auto Scaling (BP7)

Otra forma de mitigar los ataques tanto a la capa de infraestructura como a la de aplicación es operar a escala. Si tiene aplicaciones web, puede usar equilibradores de carga para distribuir el tráfico entre varias instancias de Amazon EC2 que estén sobreaprovisionadas o configuradas para escalarse automáticamente. Estas instancias pueden gestionar aumentos repentinos de tráfico que se producen por cualquier motivo, incluido un ataque de multitudes flash o DDoS en la capa de aplicación. Puede configurar alarmas de Amazon CloudWatch para iniciar Auto Scaling para escalar automáticamente el tamaño de su Amazon EC2 Fleet en respuesta a los eventos que defina, como CPU, RAM, E/S de red e incluso métricas personalizadas. Este enfoque protege la disponibilidad de la aplicación cuando hay un aumento inesperado del volumen de solicitudes. Cuando se utiliza Amazon CloudFront, Application Load Balancer, Classic Load Balancers o Network Load Balancer con la aplicación, la distribución (Amazon CloudFront) o el equilibrador de carga se encargan de la negociación de TLS. Estas características ayudan a proteger las instancias del impacto de los ataques basados en TLS al escalar para gestionar las solicitudes legítimas y los ataques de abuso de TLS.

Para obtener más información sobre el uso de Amazon CloudWatch para invocar Auto Scaling, consulte el tema Monitoreo de las métricas de CloudWatch para los grupos e instancias de Auto Scaling.

Amazon EC2 proporciona capacidad de computación de tamaño variable para que pueda escalarla vertical u horizontalmente de forma rápida a medida que cambien los requisitos. Para escalar

horizontalmente debe agregar instancias automáticamente a su aplicación mediante el escalado del tamaño de su grupo de Amazon EC2 Auto Scaling. Puede escalar verticalmente mediante el uso de tipos de instancias de EC2 más grandes.

Elastic Load Balancing (BP6)

Los ataques DDoS de gran envergadura pueden superar la capacidad de una sola instancia de Amazon EC2. Con Elastic Load Balancing (ELB), puede reducir el riesgo de sobrecargar la aplicación al distribuir el tráfico entre muchas instancias de backend. Elastic Load Balancing puede escalarse automáticamente, lo que le permite administrar volúmenes más grandes si tiene tráfico adicional imprevisto, por ejemplo, debido a multitudes flash o ataques DDoS. Para las aplicaciones creadas en una Amazon VPC, hay tres tipos de ELB que se deben tener en cuenta, según el tipo de aplicación: Application Load Balancer (ALB), Classic Load Balancer (CLB) y Network Load Balancer (NLB).

Para las aplicaciones web, puede usar el Application Load Balancer para dirigir el tráfico en función del contenido y aceptar solo solicitudes web bien formadas. El Application Load Balancer bloquea muchos ataques DDoS comunes, como inundaciones de SYN o ataques de reflexión UDP, lo que protege su aplicación del ataque. El Application Load Balancer se escala automáticamente para absorber el tráfico adicional cuando se detectan estos tipos de ataques. Las actividades de escalado debido a los ataques a la capa de infraestructura son transparentes para los clientes de AWS y no inciden en la factura.

Para obtener más información sobre la protección de aplicaciones web con Application Load Balancer, consulte el tema Introducción a Application Load Balancers.

Para las aplicaciones basadas en TCP, puede utilizar el Network Load Balancer para dirigir el tráfico hacia los destinos (por ejemplo, instancias de Amazon EC2) con una latencia ultrabaja. Una consideración clave del Network Load Balancer es que todo el tráfico que llegue al equilibrador de carga en un agente de escucha válido se dirigirá hacia sus destinos, no se absorberá. Puede usar Shield Advanced para configurar la protección contra DDoS para las direcciones IP elásticas. Cuando se asigna una dirección IP elástica por zona de disponibilidad al Network Load Balancer, Shield Advanced aplica las protecciones DDoS relevantes para el tráfico del Network Load Balancer.

Para obtener más información sobre la protección de aplicaciones TCP con el Network Load Balancer, consulte el tema Introducción a los Network Load Balancers

Beneficios de las ubicaciones de borde de AWS para escalar (BP1, BP3)

El acceso a conexiones de Internet diversas y muy escaladas puede aumentar significativamente su capacidad para optimizar la latencia y el rendimiento de los usuarios, absorber los ataques DDoS y aislar los errores a la vez que permite minimizar el impacto de la disponibilidad de su aplicación. Las ubicaciones de borde de AWS proporcionan una capa adicional de infraestructura de red que ofrece estos beneficios a cualquier aplicación web que utilice Amazon CloudFront, Global Accelerator y Amazon Route 53. Estos servicios le permiten proteger de manera integral en el borde las aplicaciones que se ejecutan desde las regiones de AWS.

Entrega de aplicaciones web en el borde (BP1)

Amazon CloudFront es un servicio que se puede utilizar para entregar todo el sitio web, incluido contenido estático, dinámico, de streaming e interactivo. Las conexiones persistentes y la configuración de período de vida (TTL, por sus siglas en inglés) variable se pueden usar para descargar el tráfico de su origen, incluso si no está sirviendo contenido que se pueda almacenar en caché. El uso de estas características de CloudFront reduce el número de solicitudes y conexiones TCP a su origen, lo que ayuda a proteger su aplicación web de las inundaciones de HTTP. CloudFront solo acepta conexiones bien formadas, lo que ayuda a evitar que muchos ataques DDoS comunes, como inundaciones de SYN y ataques de reflexión UDP, lleguen a su origen. Además, los ataques DDoS se aíslan geográficamente cerca del origen, lo que impide que el tráfico afecte a otras ubicaciones. Estas capacidades pueden mejorar significativamente su capacidad de seguir sirviendo tráfico a los usuarios durante ataques DDoS a gran escala. Puede usar CloudFront para proteger un origen en AWS o en cualquier otro lugar de Internet.

Si utiliza Amazon S3 para ofrecer contenido estático en Internet, AWS recomienda que utilice Amazon CloudFront para proteger su bucket. Puede usar la identificación de acceso de origen (OAI, por sus siglas en inglés) para garantizar que los usuarios solo accedan a los objetos mediante las URL de CloudFront.

Para obtener más información sobre la OAI, consulte el tema Restricción del acceso a contenido de Amazon S3 mediante una identidad de acceso de origen.

Para obtener más información sobre cómo proteger y optimizar el rendimiento de las aplicaciones web con Amazon CloudFront, consulte el tema Introducción a CloudFront.

Protección del tráfico de red más lejos del origen con AWS Global Accelerator (BP1)

Global Accelerator es un servicio de red que mejora la disponibilidad y el rendimiento del tráfico de los usuarios hasta en un 60 %. Esto se logra ingresando el tráfico en la ubicación de borde más cercana a los usuarios y enrutándolo a través de la infraestructura de red global de AWS a su aplicación, tanto si se ejecuta en una como en varias regiones de AWS.

Global Accelerator dirige el tráfico TCP y UDP al punto de conexión óptimo en función del rendimiento en la región de AWS más cercana al usuario. Si se produce un error en la aplicación, Global Accelerator proporciona conmutación por error al siguiente mejor punto de conexión en 30 segundos. Para proteger las aplicaciones, Global Accelerator utiliza la gran capacidad de la red global de AWS y las integraciones con Shield, como la capacidad proxy de SYN sin estado que desafía los nuevos intentos de conexión y solo sirve a los usuarios finales legítimos.

Puede implementar una arquitectura resistente a DDoS que proporcione muchos de los mismos beneficios que las prácticas recomendadas de entrega de aplicaciones web en el borde, incluso si su aplicación utiliza protocolos no compatibles con CloudFront o si está operando una aplicación web que requiere direcciones IP estáticas globales. Por ejemplo, puede requerir direcciones IP que los usuarios finales puedan agregar a la lista de permitidas en sus firewalls y que ningún otro cliente de AWS las utilice. En estos escenarios, puede usar Global Accelerator para proteger las aplicaciones web que se ejecutan en Application Load Balancer y, junto con AWS WAF, detectar y mitigar las inundaciones de solicitudes en la capa de aplicaciones web.

Para obtener más información sobre cómo proteger y optimizar el rendimiento del tráfico de red mediante Global Accelerator, consulte el tema Introducción a AWS Global Accelerator.

Resolución de nombres de dominio en el borde (BP3)

Amazon Route 53 es un servicio de sistema de nombres de dominio (DNS, por sus siglas en inglés) escalable y de alta disponibilidad que se puede utilizar para dirigir el tráfico a su aplicación web. Incluye características avanzadas como el flujo de tráfico, las comprobaciones de estado y la supervisión, el enrutamiento basado en la latencia y el DNS geográfico. Estas características avanzadas le permiten controlar cómo responde el servicio a las solicitudes de DNS para mejorar el rendimiento de su aplicación web y evitar interrupciones del sitio.

Amazon Route 53 utiliza técnicas como la fragmentación aleatoria y la creación de bandas anycast, que pueden ayudar a los usuarios a acceder a su aplicación incluso si el servicio de DNS es el objetivo de un ataque DDoS.

Con la fragmentación aleatoria, cada servidor de nombres de su conjunto de delegación corresponde a un conjunto único de ubicaciones de borde y rutas de Internet. Esto proporciona una mayor tolerancia a errores y minimiza la superposición entre los clientes. Si un servidor de nombres del conjunto de delegación no está disponible, los usuarios pueden volver a intentarlo y recibir una respuesta de otro servidor de nombres en una ubicación de borde diferente.

Las bandas anycast permiten que cada solicitud de DNS sea atendida por la ubicación más óptima, lo que dispersa la carga de la red y reduce la latencia del DNS. Así se logra dar una respuesta más rápida a los usuarios. Además, Amazon Route 53 puede detectar anomalías en el origen y el volumen de las consultas de DNS, y priorizar las solicitudes de los usuarios que se sabe que son fiables.

Para obtener más información sobre el uso de Amazon Route 53 para dirigir a los usuarios hacia su aplicación, consulte el tema Introducción a Amazon Route 53.

Defensa de la capa de aplicación (BP1, BP2)

Muchas de las técnicas analizadas hasta ahora en este documento son eficaces a la hora de mitigar el impacto de los ataques DDoS a la capa de infraestructura en la disponibilidad de su aplicación. Para defenderse también de los ataques a la capa de aplicación, debe implementar una arquitectura que le permita detectar, escalar para absorber y bloquear solicitudes malintencionadas de forma específica. Esta es una consideración importante porque los sistemas de mitigación de DDoS basados en red generalmente son ineficaces a la hora de mitigar ataques complejos en la capa de aplicación.

Detección y filtrado de solicitudes web malintencionadas (BP1, BP2)

Cuando su aplicación se ejecuta en AWS, puede aprovechar tanto Amazon CloudFront como AWS WAF para ayudar a defenderse de los ataques DDoS en la capa de aplicación.

Amazon CloudFront le permite almacenar en caché contenido estático y servirlo desde ubicaciones de borde de AWS, lo que puede ayudar a reducir la carga en el origen. También puede ayudar a reducir la carga del servidor al evitar que el tráfico no web llegue a su origen. Además, CloudFront puede cerrar automáticamente las conexiones de atacantes de lectura lenta o escritura lenta (por ejemplo, Slowloris).

AWS WAF le permite configurar listas de control de acceso web (ACL web) en sus distribuciones de CloudFront o Application Load Balancers para filtrar y bloquear solicitudes en función de las

firmas de solicitud. Cada ACL web consta de reglas que puede configurar para que coincidan con cadenas o expresiones regulares con uno o más atributos de solicitud, como el identificador uniforme de recursos (URI), la cadena de consulta, el método HTTP o la clave de encabezado. Además, al usar las reglas basadas en tasas de AWS WAF, puede bloquear automáticamente las direcciones IP de los actores maliciosos cuando las solicitudes que coincidan con una regla excedan un umbral definido por usted.

Las solicitudes de direcciones IP de clientes infractores recibirán respuestas de error 403 Forbidden y permanecerán bloqueadas hasta que las tasas de solicitudes caigan por debajo del umbral. Esto es útil para mitigar los ataques de inundación de HTTP que se disfrazan de tráfico web normal. Para bloquear los ataques en función de la reputación de la dirección IP, puede crear reglas mediante condiciones de coincidencia de IP o utilizar reglas administradas para las AWS WAF que ofrecen los vendedores en AWS Marketplace. AWS WAF ofrece directamente AWS Managed Rules como un servicio administrado en el que puede elegir grupos de reglas de reputación de IP. El grupo de reglas de la lista de reputación de IP de Amazon contiene reglas basadas en la inteligencia de amenazas interna de Amazon. Es útil si quiere bloquear direcciones IP normalmente asociadas a bots u otras amenazas. El grupo de reglas de lista de IP anónimas contiene reglas para bloquear las solicitudes de los servicios que permiten ocultar la identidad del lector. Entre estas se incluyen solicitudes de la VPN, proxies, nodos Tor y plataformas en la nube (incluido AWS). Tanto AWS WAF como CloudFront también le permiten establecer restricciones geográficas para bloquear o permitir solicitudes de países seleccionados. Esto ayuda a bloquear los ataques desde ubicaciones geográficas en las que no espera servir a los usuarios.

Para ayudar a identificar solicitudes malintencionadas, revise los registros de su servidor web o use las características de registro y solicitudes de muestra de AWS WAF. Al habilitar el registro de AWS WAF, obtiene información detallada sobre el tráfico analizado por la ACL web. AWS WAF permite filtrar registros, por lo que puede especificar qué solicitudes web se registran y qué solicitudes se descartan del registro después de la inspección.

La información registrada en los registros incluye la hora en que AWS WAF recibió la solicitud de su recurso de AWS, información detallada sobre la solicitud y la acción coincidente para cada regla solicitada. Las solicitudes de muestra proporcionan detalles sobre las solicitudes de las últimas tres horas que coincidieron con una de sus reglas de AWS WAF. Puede usar esta información para identificar firmas de tráfico potencialmente malintencionado y crear una nueva regla para denegar esas solicitudes. Si ve varias solicitudes con una cadena de consulta aleatoria, asegúrese de permitir solo los parámetros de cadena de consulta que son relevantes para almacenar en caché la aplicación. Esta técnica es útil para mitigar un ataque de eliminación de caché contra su origen.

Si está suscrito a AWS Shield Advanced, puede contratar al equipo de respuesta de AWS Shield (SRT) para que lo ayude a crear reglas para mitigar un ataque que esté afectando a la disponibilidad de su aplicación. Puede otorgar acceso limitado a AWS SRT a Shield Advanced y a las API de AWS WAF. AWS SRT solo accederá a estas API para aplicar mitigaciones a su cuenta si usted lo autoriza de forma explícita. Para obtener más información, consulte la sección Soporte de este documento.

Puede utilizar AWS Firewall Manager para configurar y administrar de forma centralizada las reglas de seguridad, como las protecciones de Shield Advanced y las reglas de AWS WAF, en toda su organización. Su cuenta de administración de AWS Organizations puede designar una cuenta de administrador, que está autorizada para crear políticas de Firewall Manager. Estas políticas le permiten definir criterios, como el tipo de recurso y las etiquetas, que determinan dónde se aplican las reglas. Esto resulta útil si tiene varias cuentas y desea estandarizar su protección.

Para obtener más información acerca de:

- Las reglas administradas de AWS para AWS WAF, consulte el tema Reglas administradas de AWS para AWS WAF.
- Usar la restricción geográfica para limitar el acceso a la distribución de CloudFront, consulte el tema Cómo restringir la distribución geográfica de su contenido.
- Usar AWS WAF, consulte los temas:
  - Introducción a AWS WAF
  - Registro de información del tráfico de la ACL web
  - Visualizar una muestra de solicitudes web
- Configurar reglas basadas en tasas, consulte el tema sobre la protección de sitios web y servicios mediante reglas basadas en tasas para AWS WAF
- Cómo administrar la implementación de reglas de AWS WAF en sus recursos de AWS con Firewall Manager, consulte los temas:
  - Introducción a las políticas de AWS WAF de Firewall Manager
  - Introducción a las políticas de Shield Advanced de Firewall Manager

Reducción de la superficie de ataque

Otra consideración importante a la hora de diseñar una solución de AWS es limitar las oportunidades que tiene un atacante de atacar su aplicación. Este concepto se conoce como reducción de la superficie de ataque. Los recursos que no están expuestos a Internet son más difíciles de atacar, lo que limita las opciones que tiene un atacante de vulnerar la disponibilidad de la aplicación.

Por ejemplo, si no quiere que los usuarios interactúen directamente con ciertos recursos, asegúrese de que no se pueda acceder a esos recursos desde Internet. Del mismo modo, no acepte tráfico de usuarios o aplicaciones externas en puertos o protocolos que no sean necesarios para la comunicación.

En la siguiente sección, AWS explica las prácticas recomendadas para guiarlo a la hora de reducir la superficie de ataque y limitar la exposición a Internet de su aplicación.

Temas

Ocultar los recursos de AWS (BP1, BP4, BP5)

Por lo general, los usuarios pueden utilizar una aplicación de forma rápida y sencilla sin necesidad de que los recursos de AWS estén completamente expuestos en Internet. Por ejemplo, si tiene instancias de Amazon EC2 detrás de un Elastic Load Balancing, es posible que las instancias en sí mismas no tengan que ser de acceso público. En cambio, puede proporcionar a los usuarios acceso a Elastic Load Balancing en ciertos puertos TCP y permitir que solo Elastic Load Balancing se comunique con las instancias. Puede hacerlo mediante la configuración de grupos de seguridad y listas de control de acceso (ACL) de red en su Amazon Virtual Private Cloud (Amazon VPC). Amazon VPC le permite aprovisionar una sección aislada de forma lógica de la nube de AWS donde puede lanzar recursos de AWS en una red virtual definida por usted.

Los grupos de seguridad y las ACL de la red son similares en el sentido de que le permiten controlar el acceso a los recursos de AWS dentro de su VPC. Sin embargo, los grupos de seguridad le permiten controlar el tráfico entrante y saliente en el nivel de la instancia, mientras que las ACL de red ofrecen capacidades similares en el nivel de subred de la VPC. No se aplican cargos adicionales por el uso de grupos de seguridad o ACL de red.

Grupos de seguridad y listas de control de acceso a la red (ACL de red) (BP5)

Puede elegir si desea especificar los grupos de seguridad al lanzar una instancia o asociar la instancia a un grupo de seguridad más adelante. Todo el tráfico procedente de Internet hacia un grupo de seguridad se deniega implícitamente a menos que cree una regla allow para permitir el tráfico. Por ejemplo, si tiene una aplicación web que utiliza un Elastic Load Balancing y varias instancias de Amazon EC2, puede decidir crear un grupo de seguridad para Elastic Load Balancing (grupo de seguridad de Elastic Load Balancing) y otro para las instancias (grupo de seguridad del servidor de aplicaciones web). A continuación, puede crear una regla allow para permitir el tráfico procedente de Internet hacia el grupo de seguridad de ELB y otra regla para permitir el tráfico procedente del grupo de seguridad de ELB hacia el grupo de seguridad del servidor de aplicaciones web. Esto garantiza que el tráfico de Internet no pueda comunicarse directamente con las instancias de Amazon EC2, lo que dificulta que un atacante obtenga información sobre su aplicación y que le afecte.

Al crear ACL de red, puede especificar reglas de permiso y denegación. Esto resulta útil en caso de que quiera denegar explícitamente determinados tipos de tráfico hacia su aplicación. Por ejemplo, puede definir las direcciones IP (como rangos de CIDR), los protocolos y los puertos de destino que no pueden acceder a toda la subred. Si su aplicación se utiliza para tráfico TCP, puede crear una regla que deniegue todo el tráfico UDP, o viceversa. Esta opción es útil al responder a ataques DDoS porque le permite crear sus propias reglas para mitigar el ataque cuando conoce las IP de origen u otra firma.

Si está suscrito a AWS Shield Advanced, puede registrar direcciones IP elásticas como recursos protegidos. Los ataques DDoS contra direcciones IP elásticas que se han registrado como recursos protegidos se detectan más rápidamente, lo que puede reducir el tiempo de mitigación. Cuando se detecta un ataque, los sistemas de mitigación de DDoS leen la ACL de red que corresponde a la IP elástica de destino y la aplican en el borde de la red de AWS. Esto reduce significativamente el riesgo de impacto de una serie de ataques DDoS en la capa de infraestructura.

Para obtener más información sobre la configuración de grupos de seguridad y las ACL de red para optimizar la resiliencia de DDoS, consulte el tema sobre cómo ayudar a prepararse para los ataques DDoS reduciendo la superficie de ataque.

Para obtener más información sobre el uso de Shield Advanced con direcciones IP elásticas como recursos protegidos, consulte los pasos para suscribirse a AWS Shield Advanced.

Protección del origen (BP1, BP5)

Si utiliza Amazon CloudFront con un origen que está dentro de su VPC, es posible que desee asegurarse de que solo su distribución de CloudFront pueda reenviar solicitudes a su origen. Con los encabezados de la solicitud de borde a origen, puede agregar o anular el valor de los encabezados de solicitud existentes cuando CloudFront reenvía las solicitudes a su origen. Puede utilizar los encabezados personalizados de origen, por ejemplo, el encabezado X-Shared-Secret, para ayudar a validar que las solicitudes realizadas a su origen se hayan enviado desde CloudFront.

Para obtener más información sobre cómo proteger su origen con encabezados personalizados de origen, consulte el tema Agregar encabezados personalizados a solicitudes de origen y Restringir el acceso a los balanceadores de carga de aplicaciones.

Para obtener una guía sobre la implementación de una solución de muestra para rotar automáticamente el valor de los encabezados personalizados de origen para restringir el acceso al origen, consulte el tema cómo mejorar la seguridad de origen de Amazon CloudFront con AWS WAF y Secrets Manager.

Como alternativa, puede usar una función de AWS Lambda para actualizar automáticamente las reglas de su grupo de seguridad para permitir solo el tráfico de CloudFront. Esto mejora la seguridad de su origen al ayudar a garantizar que los usuarios malintencionados no puedan eludir CloudFront y AWS WAF al acceder a su aplicación web.

Para obtener más información sobre cómo proteger su origen mediante la actualización automática de los grupos de seguridad, consulte el encabezado X-Shared-Secret y el tema sobre cómo actualizar automáticamente sus grupos de seguridad para Amazon CloudFront y AWS WAF con AWS Lambda.

Protección de los puntos de conexión de las API (BP4)

Por lo general, al exponer una API al público, existe el riesgo de que el frontend de la API pueda sufrir un ataque DDoS. Para ayudar a reducir este riesgo, puede utilizar Amazon API Gateway como entrada a las aplicaciones que se ejecutan en Amazon EC2, AWS Lambda o en otro lugar. Al usar Amazon API Gateway, no necesita sus propios servidores para el frontend de la API y puede ofuscar otros componentes de su aplicación. Al dificultar la detección de los componentes de su aplicación, ayuda a evitar que dichos recursos de AWS sufran ataques DDoS.

Al utilizar Amazon API Gateway, puede elegir entre dos tipos de puntos de conexión de API. La primera es la opción predeterminada: puntos de conexión de API optimizados para el borde a los que

se accede a través de una distribución de Amazon CloudFront. Sin embargo, API Gateway es quién crea y administra la distribución, por lo que usted no tiene ningún control sobre ella. La segunda opción es usar un punto de conexión de API regional al que se accede desde la misma región de AWS en la que se ha implementado la API de REST. AWS recomienda utilizar el segundo tipo de punto de conexión y que lo asocie a su propia distribución de Amazon CloudFront. Esto le da control sobre la distribución de Amazon CloudFront y la capacidad de usar AWS WAF para así proteger de la capa de aplicación. Este modo le proporciona acceso a una capacidad de mitigación de DDoS escalada en toda la red global de borde de AWS.

Al usar Amazon CloudFront y AWS WAF con Amazon API Gateway, deberá configurar las siguientes opciones:

- Parametrizar el comportamiento de la memoria caché para que sus distribuciones reenvíen todos los encabezados al punto de conexión regional de API Gateway. De este modo, CloudFront tratará el contenido como dinámico y omitirá el almacenamiento en caché del contenido.
- Proteger la API Gateway de todo acceso directo configurando la distribución para que incluya el encabezado personalizado de origen `x-api-key` y estableciendo el valor de la clave API en API Gateway.
- Proteger el backend del exceso de tráfico configurando límites de velocidad estándar o de ráfaga para cada método en sus API de REST.

Para obtener más información sobre la creación de API con Amazon API Gateway, consulte el tema Introducción a Amazon API Gateway.

Técnicas operativas

Las técnicas de mitigación de este documento le ayudan a diseñar aplicaciones que son inherentemente resistentes a los ataques DDoS. En muchos casos, también es útil saber cuándo un ataque DDoS tiene como objetivo su aplicación para que pueda tomar medidas de mitigación. En esta sección se explican las prácticas recomendadas para obtener visibilidad del comportamiento anormal, las alertas y la automatización, la administración de la protección a escala y la participación de AWS para obtener soporte adicional.

Temas

Visibilidad

Cuando una métrica operativa clave se desvía sustancialmente del valor esperado, es posible que un atacante esté intentando apuntar hacia la disponibilidad de la aplicación. Familiarizarse con el comportamiento normal de su aplicación significa que puede actuar con mayor rapidez cuando detecte una anomalía. Amazon CloudWatch puede ayudarle mediante la supervisión de las aplicaciones en las que se ejecuta AWS. Por ejemplo, puede recopilar y realizar el seguimiento de métricas, recopilar y supervisar archivos de registro, establecer alarmas y reaccionar automáticamente a los cambios en sus recursos de AWS.

Si sigue la arquitectura de referencia resistente a DDoS al diseñar su aplicación, los ataques comunes a la capa de infraestructura se bloquearán antes de llegar a su aplicación. Si está suscrito a AWS Shield Advanced, tendrá acceso a una serie de métricas de CloudWatch que pueden indicar que su aplicación es un objetivo. Por ejemplo, puede configurar alarmas para que le notifiquen cuando haya un ataque DDoS en curso, de modo que pueda comprobar el estado de su aplicación y decidir si desea activar AWS SRT. Puede configurar la métrica de `DDoSDetected` para que le indique si se ha detectado un ataque. Si quiere recibir alertas en función del volumen de ataques, también puede usar las métricas `DDoSAttackBitsPerSecond`, `DDoSAttackPacketsPerSecond` o `DDoSAttackRequestsPerSecond`. Puede supervisar estas métricas integrando CloudWatch con sus propias herramientas o utilizando herramientas proporcionadas por terceros, como Slack o PagerDuty.

Un ataque a la capa de aplicación puede elevar muchas métricas de Amazon CloudWatch. Si usa AWS WAF, puede utilizar CloudWatch para supervisar y activar alarmas en caso de aumento de las solicitudes que haya configurado para que se permitan, cuenten o bloqueen en AWS WAF. De este modo, recibirá una notificación si el nivel de tráfico supera lo que su aplicación puede gestionar. También puede utilizar Amazon CloudFront, Amazon Route 53, Application Load Balancer, Network Load Balancer, Amazon EC2 y métricas de Auto Scaling que se rastrean en CloudWatch para detectar cambios que puedan indicar un ataque DDoS.

En la tabla Métricas recomendadas de Amazon CloudWatch se enumeran las descripciones de las métricas de CloudWatch que se utilizan habitualmente para detectar y reaccionar a los ataques DDoS.

Tabla 3 - Métricas recomendadas de Amazon CloudWatch

Tema	Métrica	Descripción
AWS Shield Advanced	DDoSDetected	Indica un evento DDoS para un nombre de recurso de Amazon (ARN) determinado.
AWS Shield Advanced	DDoSAttackBitsPerSecond	Número de bytes observado durante un evento DDoS de un ARN específico. Esta métrica solo está disponible para los eventos DDoS de las capas 3 y 4.
AWS Shield Advanced	DDoSAttackPacketsPerSecond	Cantidad de paquetes observados durante un evento DDoS de un ARN específico. Esta métrica solo está disponible para los eventos DDoS de las capas 3 y 4.
AWS Shield Advanced	DDoSAttackRequestsPerSecond	Cantidad de solicitudes observadas durante un evento DDoS de un ARN específico. Esta métrica solo está

Tema	Métrica	Descripción
		disponible para los eventos DDoS de la capa 7 y solo figurará en los informes de los eventos de la capa 7 más importantes.
AWS WAF	AllowedRequests	Número de solicitudes web permitidas.
AWS WAF	BlockedRequests	Número de solicitudes web bloqueadas.
AWS WAF	CountedRequests	Número de solicitudes web contabilizadas.
AWS WAF	PassedRequests	Número de solicitudes aprobadas. Solo se usa para solicitudes que pasan por una evaluación de grupo de reglas sin coincidir con ninguna de las reglas del grupo de reglas.
Amazon CloudFront	Requests	Número de solicitudes HTTP/S.
Amazon CloudFront	TotalErrorRate	Porcentaje de todas las solicitudes para las que el estado de código HTTP es 4xx o 5xx.
Amazon Route 53	HealthCheckStatus	Estado del punto de conexión de la comprobación de estado.

Tema	Métrica	Descripción
Application Load Balancer	ActiveConnectionCount	Número total de conexiones TCP simultáneas activas desde los clientes al equilibrador de carga y desde el equilibrador de carga a los destinos.
Application Load Balancer	ConsumedLCUs	Número de unidades de capacidad del equilibrador de carga (LCU) usadas por el equilibrador de carga.
Application Load Balancer	HTTPCode_ELB_4XX_Count HTTPCode_ELB_5XX_Count	Número de códigos de error del cliente HTTP 4xx o 5xx generados por el equilibrador de carga.
Application Load Balancer	NewConnectionCount	Número total de conexiones TCP nuevas establecidas desde los clientes al equilibrador de carga y desde el equilibrador de carga a los destinos.
Application Load Balancer	ProcessedBytes	Número total de bytes procesados por el equilibrador de carga.
Application Load Balancer	RejectedConnectionCount	Número de conexiones que se rechazaron porque el equilibrador de carga alcanzó el número máximo de conexiones.
Application Load Balancer	RequestCount	Cantidad de solicitudes procesadas.

Tema	Métrica	Descripción
Application Load Balancer	TargetConnectionErrorCount	Número de conexiones que no se establecieron correctamente entre el equilibrador de carga y el destino.
Application Load Balancer	TargetResponseTime	Tiempo transcurrido, en segundos, desde que la solicitud ha salido del equilibrador de carga hasta que se recibe una respuesta del destino.
Application Load Balancer	UnHealthyHostCount	Número de destinos que se considera que no está en buen estado.
Network Load Balancer	ActiveFlowCount	Número total de flujos (o conexiones) TCP simultáneos de clientes a destinos.
Network Load Balancer	ConsumedLCUs	Número de unidades de capacidad del equilibrador de carga (LCU) usadas por el equilibrador de carga.
Network Load Balancer	NewFlowCount	Número total de flujos (o conexiones) TCP nuevos establecidos desde los clientes a los destinos en el periodo indicado.
Network Load Balancer	ProcessedBytes	Número total de bytes procesados por el equilibrador de carga, incluidos los encabezados TCP/IP.

Tema	Métrica	Descripción
Global Accelerator	NewFlowCount	Número total de flujos (o conexiones) TCP nuevos establecidos desde los clientes a los puntos de conexión en el periodo indicado.
Global Accelerator	ProcessedBytesIn	Número total de bytes entrantes procesados por el acelerador, incluidos los encabezados TCP/IP.
Auto Scaling	GroupMaxSize	Tamaño máximo del grupo de Auto Scaling.
Amazon EC2	CPUUtilization	Porcentaje de unidades de computación EC2 asignadas en uso actualmente en la instancia.
Amazon EC2	NetworkIn	Número de bytes recibidos por la instancia en todas las interfaces de red.

Para obtener más información sobre el uso de Amazon CloudWatch para detectar ataques DDoS en su aplicación, consulte el tema Introducción a Amazon CloudWatch.

Para ver un ejemplo de un panel creado con algunas de las métricas de la tabla anterior, consulte el tema sobre un sistema de supervisión de base de referencia personalizada.

AWS incluye varias métricas y alarmas adicionales para notificarle sobre un ataque y ayudarlo a supervisar los recursos de su aplicación. La consola de AWS Shield o la API proporcionan un resumen de eventos por cuenta y detalles sobre los ataques que se han detectado.

Además, el panel del entorno de amenazas globales proporciona información resumida sobre todos los ataques DDoS detectados por AWS. Esta información puede resultar útil para comprender mejor las amenazas DDoS en una masa más amplia de aplicaciones, además de las tendencias en los ataques, y también para compararlas con los ataques que puede haber observado.

Si está suscrito a AWS Shield Advanced, el panel de servicio muestra métricas de detección y mitigación adicionales y detalles del tráfico de red para los eventos detectados en los recursos protegidos. AWS Shield evalúa el tráfico a su recurso protegido en múltiples dimensiones. Cuando se detecta una anomalía, AWS Shield crea un evento e informa de la dimensión del tráfico en la que se observó la anomalía. Con una mitigación colocada, esto protege a su recurso de recibir tráfico excesivo y tráfico que coincida con una firma de evento DDoS conocida.

Las métricas de detección se basan en flujos de red de muestra o en registros de AWS WAF cuando se asocia una ACL web con el recurso protegido. Las métricas de mitigación se basan en el tráfico observado por los sistemas de mitigación de DDoS de Shield. Las métricas de mitigación son una medición más precisa del tráfico en su recurso.

La métrica de los principales contribuyentes de la red proporciona información sobre el origen del tráfico durante un evento detectado. Puede ver los contribuyentes de mayor volumen y ordenarlos por elementos como el protocolo, el puerto de origen y los indicadores TCP. La métrica de los principales contribuyentes incluye métricas para todo el tráfico observado en el recurso en varias dimensiones. Proporciona dimensiones métricas adicionales que puede usar para comprender el tráfico de red que se envía a su recurso durante un evento.

Esto también incluye detalles sobre las acciones realizadas automáticamente para mitigar los ataques de DDoS. Esta información facilita la investigación de anomalías, la exploración de las dimensiones del tráfico y comprender mejor las medidas adoptadas por Shield Advanced para proteger su disponibilidad.

Otra herramienta que puede ayudarlo a obtener visibilidad del tráfico dirigido a su aplicación son los registros de flujo de VPC. En una red tradicional, puede usar registros de flujo de red para solucionar problemas de conectividad y seguridad y para asegurarse de que las reglas de acceso a la red funcionan como se espera. Mediante el uso de registros de flujo de la VPC, puede capturar información sobre el tráfico IP que va y viene de las interfaces de red de la VPC.

Cada registro de flujo incluye lo siguiente: direcciones IP de origen y destino, puertos de origen y destino, protocolo y el número de paquetes y bytes transferidos durante la ventana de captura. Puede utilizar esta información para ayudar a identificar anomalías en el tráfico de red y para identificar un vector de ataque específico. Por ejemplo, la mayoría de los ataques de reflexión UDP

tienen puertos de origen específicos, como el puerto de origen 53 para la reflexión de DNS. Se trata de una firma de ataque clara que puede identificar en el registro de flujo. En respuesta, puede elegir bloquear el puerto de origen específico en el nivel de instancia o crear una regla de ACL de red para bloquear todo el protocolo si su aplicación no lo requiere.

Para obtener más información sobre el uso de registros de flujo de VPC para identificar anomalías de red y vectores de ataques DDoS, consulte Logs de flujo de VPC y el tema sobre registros de flujo de VPC y cómo registrar y ver flujos de tráfico de red.

Gestión de la visibilidad y la protección en varias cuentas

En situaciones en las que opera en varias cuentas de AWS y tiene varios componentes que proteger, el uso de técnicas que le permitan operar a escala y reducir la sobrecarga operativa aumenta sus capacidades de mitigación. Al administrar recursos protegidos de AWS Shield Advanced en varias cuentas, puede configurar una supervisión centralizada mediante AWS Firewall Manager y AWS Security Hub. Con Firewall Manager, puede crear una política de seguridad que imponga el cumplimiento de la protección DDoS en todas sus cuentas. Puede usar estos dos servicios juntos para administrar sus recursos protegidos en varias cuentas y centralizar la supervisión de esos recursos.

Security Hub se integra automáticamente con Firewall Manager, lo que permite a los clientes de Shield Advanced ver los hallazgos de seguridad en un único panel, junto con otras alertas de seguridad de alta prioridad y estados de cumplimiento. Por ejemplo, cuando Shield Advanced detecta tráfico anómalo destinado a un recurso protegido en cualquier cuenta de AWS dentro del ámbito, este hallazgo se verá en la consola de Security Hub. Si se configura, Firewall Manager puede hacer que el recurso cumpla los requisitos automáticamente al crearlo como un recurso protegido con Shield Advanced y, a continuación, actualizar Security Hub cuando el recurso tenga el estado conforme.

Para obtener más información sobre la supervisión centralizada de los recursos protegidos de Shield, consulte el tema sobre la configuración de la supervisión centralizada de los eventos de DDoS y la corrección automática de los recursos no conformes.

Soporte

Si sufre un ataque, también puede beneficiarse del soporte de AWS para evaluar la amenaza y revisar la arquitectura de su aplicación, o puede solicitar otro tipo de ayuda. Es importante crear

un plan de respuesta para los ataques DDoS antes de que ocurran. Las prácticas recomendadas descritas en este documento pretenden ser medidas proactivas que se implementan antes de lanzar una aplicación, pero es posible que se sigan produciendo ataques DDoS contra la aplicación. Revise las opciones de esta sección para determinar los recursos de soporte que mejor se adapten a su situación. Su equipo de cuentas puede evaluar su caso de uso y su aplicación, y ayudarlo con las preguntas o desafíos específicos que tenga.

Si ejecuta cargas de trabajo de producción en AWS, plantéese suscribirse a Business Support, que proporciona acceso ininterrumpido a ingenieros de soporte en la nube que pueden ayudarlo en caso de problemas con ataques DDoS. Si ejecuta cargas de trabajo de misión crítica, considere suscribirse a Enterprise Support, que proporciona la capacidad de abrir casos críticos y recibir respuestas muy rápidas de un ingeniero sénior de soporte en la nube.

Si está suscrito a AWS Shield Advanced y también a Business Support o Enterprise Support, puede configurar la interacción proactiva de Shield. Este le permite configurar comprobaciones de estado, asociarse a sus recursos y proporcionar información de contacto de operaciones de forma ininterrumpida. Cuando Shield detecte signos de DDoS y las comprobaciones de estado de sus aplicaciones muestren signos de degradación, AWS SRT se pondrá en contacto con usted de forma proactiva. Este es nuestro modelo de interacción recomendado porque permite tiempos de respuesta de AWS SRT más rápidos y permite a AWS SRT comenzar a solucionar problemas incluso antes de que se haya establecido contacto con usted.

La función de interacción proactiva requiere que se configure una comprobación de estado de Route 53 que mida con precisión el estado de su aplicación y que esté asociada al recurso protegido por Shield Advanced. Después de asociar una comprobación de estado de Route 53 a la consola de Shield, el sistema de detección de Shield Advanced utiliza el estado de la comprobación de estado como un indicador del estado de la aplicación. La característica de detección basada en el estado de Shield Advanced garantizará que se le notifique y que las mitigaciones se realicen más rápidamente cuando su aplicación no esté en buen estado. AWS SRT se pondrá en contacto con usted para buscar una solución, tanto si la aplicación en mal estado está siendo atacada por DDoS como para realizar mitigaciones adicionales según sea necesario.

Completar la configuración de la interacción proactiva incluye agregar detalles de contacto en la consola de Shield. AWS SRT utilizará esta información para ponerse en contacto con usted. Puede configurar hasta 10 contactos e incluir notas adicionales si tiene requisitos o preferencias de contacto específicos. Los contactos de interacción proactiva deben tener una función ininterrumpida como, por ejemplo, un centro de operaciones de seguridad o una persona que esté disponible de inmediato.

Puede habilitar la interacción proactiva para todos los recursos o para recursos de producción clave seleccionados donde el tiempo de respuesta es fundamental. Esto se logra asignando comprobaciones de estado únicamente a esos recursos.

También puede derivar a AWS SRT creando un caso de AWS Support mediante la consola de AWS Support o la API de Support si tiene un evento relacionado con DDoS que afecte a la disponibilidad de su aplicación.

Conclusión

Las prácticas recomendadas descritas en este documento son una guía para crear una arquitectura resistente a ataques DDoS que proteja la disponibilidad de su aplicación pues puede prevenir muchos ataques DDoS comunes en la capa de aplicación y de infraestructura. La medida en que siga estas prácticas recomendadas al diseñar su aplicación influirá en el tipo, el vector y el volumen de los ataques DDoS que pueda mitigar. Puede incorporar resiliencia sin suscribirse a un servicio de mitigación de ataques DDoS. Al elegir suscribirse a AWS Shield Advanced obtendrá características adicionales de soporte, visibilidad, mitigación y garantía de costes que protegerán aún más una arquitectura de aplicaciones ya resiliente de por sí.

Colaboradores

Entre los colaboradores de este documento, están las siguientes personas:

- Jeffrey Lyon, protección perimetral de AWS
- Rodrigo Ferroni, especialista en seguridad de AWS, TAM
- Dmitriy Novikov, arquitecto de soluciones de AWS
- Achraf Souk, arquitecto de soluciones de AWS
- Yoshihisa Nakatani, arquitecto de soluciones de AWS

Recursos

Documentación adicional:

- Best Practices for DDoS Mitigation on AWS
- Guidelines for Implementing AWS WAF
- SID324 – re:Invent 2017: Automating DDoS Response in the Cloud
- CTD304 – re:Invent 2017: Dow Jones & Wall Street Journal's Journey to Manage Traffic Spikes While Mitigating DDoS & Application Layer Threats
- CTD310 – re:Invent 2017: Living on the Edge, It's Safer Than You Think! Building Strong with Amazon CloudFront, AWS Shield, and AWS WAF
- SEC407 - re:Invent 2019: A defense-in-depth approach to building web applications
- SEC321 - re:Invent 2020: Get ahead of the curve with DDoS Response Team escalations
- William Hill: High-performance DDOS Protection with AWS

Revisiones del documento

Para recibir notificaciones sobre las actualizaciones de este documento técnico, suscríbase a la fuente RSS.

update-history-change	update-history-description	update-history-date
Documento técnico actualizado	Se ha actualizado para incluir las recomendaciones y características más recientes . AWS Global Accelerator se añade como parte de la protección integral en el borde. AWS Firewall Manager para la supervisión centralizada de los eventos de DDoS y la corrección automática de los recursos no conformes.	21 de septiembre de 2021
Documento técnico actualizado	Se ha actualizado para aclarar la eliminación de caché en la sección Detección y filtrado de solicitudes web malintencionadas (BP1, BP2) y el uso de ELB y ALB en la sección Escalar para absorber (BP6). Se han actualizado los diagramas y la tabla 2. Se ha marcado Elección de región como BP8. Se ha actualizado la sección BP7 con más detalles.	18 de diciembre de 2019
Documento técnico actualizado	Se ha actualizado para incluir el registro de AWS WAF como práctica recomendada.	1 de diciembre de 2018

Documento técnico actualizado	Se ha actualizado para incluir AWS Shield, las características de AWS WAF, AWS Firewall Manager y prácticas recomendadas relacionadas.	1 de junio de 2018
Documento técnico actualizado	Se han agregado directrices de arquitectura prescriptivas y se han actualizado para incluir AWS WAF.	1 de junio de 2016
Publicación inicial	Documento técnico publicado.	1 de junio de 2015

Avisos

Los clientes son responsables de realizar sus propias evaluaciones de la información contenida en este documento. Este documento: (a) solo tiene fines informativos, (b) representa las prácticas y las ofertas de productos vigentes de AWS, que están sujetas a cambios sin previo aviso, y (c) no crea ningún compromiso ni garantía de AWS y sus empresas afiliadas, proveedores o concesionarios de licencias. Los productos o servicios de AWS se proporcionan “tal cual”, sin garantías, representaciones ni condiciones de ningún tipo, ya sean explícitas o implícitas. Las responsabilidades y obligaciones de AWS en relación con sus clientes se rigen por los acuerdos de AWS, y este documento no modifica ni forma parte de ningún acuerdo entre AWS y sus clientes.