Whitepaper da AWS

Práticas recomendadas para resiliência contra DDoS da AWS

Práticas recomendadas para resiliência contra DDoS da AWS: Whitepaper da AWS

Table of Contents

Práticas recomendadas da AWS para resiliência contra DDoS

Data de publicação: 21 de setembro de 2021 (Revisões do documento)

Resumo

É importante proteger sua empresa contra o impacto dos ataques de negação de serviço distribuída (DDoS), bem como de outros ataques cibernéticos. Manter a confiança do cliente em seu serviço mantendo a disponibilidade e a capacidade de resposta da sua aplicação é de alta prioridade. Você também deseja evitar custos diretos desnecessários quando sua infraestrutura precisa reduzir a escala na horizontal em resposta a um ataque. A Amazon Web Services (AWS) tem o compromisso de fornecer a você as ferramentas, as práticas recomendadas e os serviços para se defender contra malfeitores na Internet. O uso dos serviços certos da AWS ajuda a garantir alta disponibilidade, segurança e resiliência.

Neste whitepaper, a AWS fornece orientações prescritivas contra DDoS para melhorar a resiliência das aplicações em execução na AWS. Isso inclui uma arquitetura de referência resistente a DDoS que pode ser usada como um guia para ajudar a proteger a disponibilidade da aplicação. Este whitepaper também descreve diferentes tipos de ataque, como ataques na camada de infraestrutura e ataques na camada de aplicação. A AWS explica quais práticas recomendadas são mais eficazes para gerenciar cada tipo de ataque. Além disso, os serviços e recursos que se encaixam em uma estratégia de mitigação de DDoS são descritos e como cada um pode ser usado para ajudar a proteger suas aplicações é explicado.

Este documento destina-se a tomadores de decisão de TI e engenheiros de segurança familiarizados com os conceitos básicos de rede, segurança e AWS. Cada seção tem links para a documentação da AWS que fornece mais detalhes sobre as práticas recomendadas ou recursos.

Introdução: Ataques de negação de serviço

Um ataque de negação de serviço (DoS) é uma tentativa deliberada de tornar um site ou aplicação indisponível para os usuários ao inundá-lo com tráfego de rede. Os invasores usam uma variedade de técnicas que consomem grandes quantidades de largura de banda da rede ou vinculam outros recursos do sistema, interrompendo o acesso de usuários legítimos. Em sua forma mais simples, um invasor solitário usa uma única fonte para realizar um ataque DoS contra um alvo, conforme mostrado na imagem a seguir.

Tabela 1: diagrama do ataque de DoS

Em um ataque de DDoS, um invasor usa várias fontes para orquestrar um ataque contra um destino. Essas fontes podem incluir grupos distribuídos de computadores infectados por malware, roteadores, dispositivos IoT e outros endpoints. O diagrama a seguir mostra que uma rede de hosts comprometidos participa do ataque, gerando uma inundação de pacotes ou solicitações para sobrecarregar o alvo.

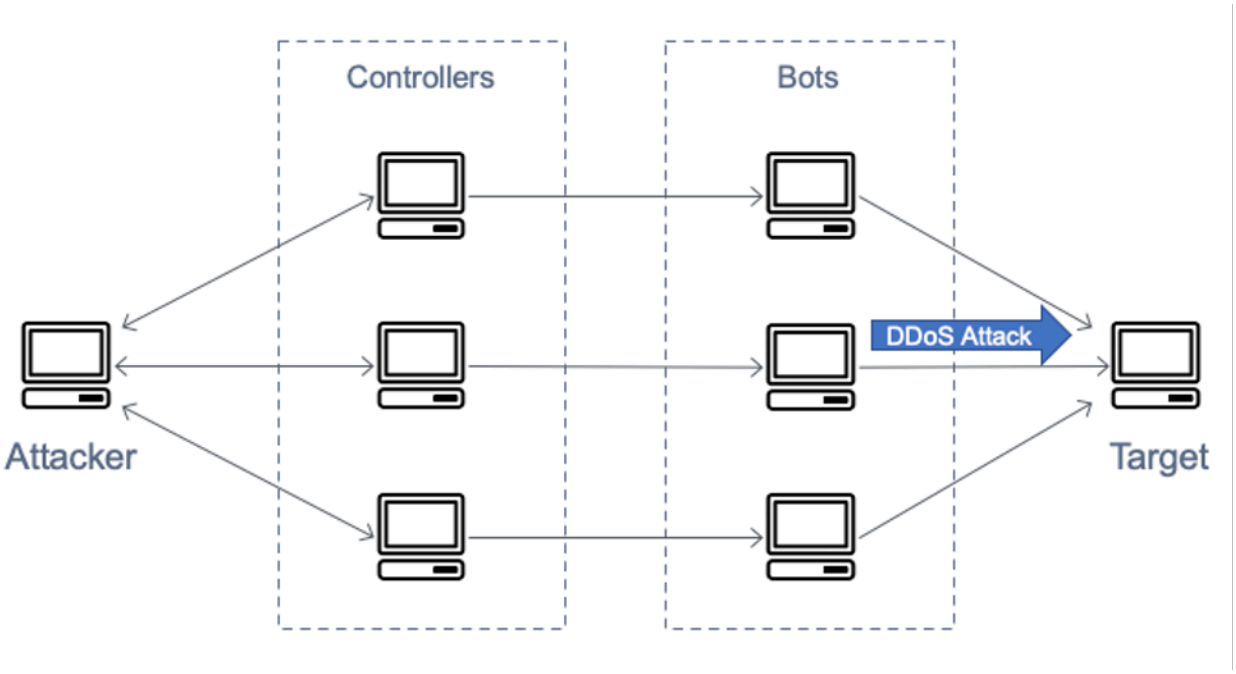

Diagrama do ataque de DDoS

Existem sete camadas no modelo de interconexão de sistemas abertos (OSI) e elas são descritas na tabela Modelo de interconexão de sistemas abertos (OSI). Os ataques de DDoS são mais comuns nas camadas três, quatro, seis e sete. Os ataques das camadas três e quatro correspondem

às camadas de Rede e Transporte do modelo OSI. Neste documento, a AWS refere-se a eles coletivamente como ataques à camada de infraestrutura. Os ataques das camadas seis e sete correspondem às camadas de Apresentação e Aplicação do modelo OSI. A AWS abordará isso em conjunto como ataques da camada de aplicação. Exemplos desses tipos de ataque serão discutidos nas seções a seguir.

Modelo de interconexão de sistemas abertos (OSI)

Nº	Camada	Unidade	Descrição	Exemplos de vetor
7	Aplicação	Dados	Processo de rede para a aplicação	Inundações HTTP, inundações de consultas de DNS
6	Apresentação	Dados	Representação de dados e criptografia	Uso abusivo de TLS
5	Sessão	Dados	Comunicação entre hosts	N/D
4	Transporte	Segmentos	Conexões e confiabilidade de ponta a ponta	Inundações SYN
3	Rede	Pacotes	Determinação do caminho e endereçamento lógico	Ataques de reflexão de UDP
2	Link de dados	Quadros	Endereçamento físico	N/D
1	Físico	Bits	Mídia, sinal e transmissão binária	N/D

Tópicos

Ataques na camada Infraestrutura

Os ataques de DDoS mais comuns, ataques de reflexão do protocolo UDP (User Datagram Protocol) e inundações de sincronização (SYN), são ataques de camada de infraestrutura. Um invasor pode usar um desses métodos para gerar grandes volumes de tráfego que podem inundar a capacidade de uma rede ou imobilizar recursos em sistemas como servidores, firewalls, sistema de prevenção de intrusões (IPS) ou balanceador de carga. Embora esses ataques possam ser fáceis de identificar, para mitigá-los de forma eficaz, você deverá ter uma rede ou sistemas que aumentem a escala verticalmente da capacidade com mais rapidez do que a inundação de tráfego de entrada. Essa capacidade extra é necessária para filtrar ou absorver o tráfego de ataque, liberando o sistema e a aplicação para responder ao tráfego legítimo do cliente.

Tópicos

Ataques de reflexão de UDP

Os ataques de reflexão do User Datagram Protocol (UDP) exploram o fato de que o UDP é um protocolo sem estado. Os invasores podem criar um pacote de solicitação UDP válido listando o endereço IP do alvo do ataque como o endereço IP de origem UDP. O invasor agora falsificou (spoof) o IP de origem do pacote de solicitação UDP. O pacote UDP contém o IP de origem falsificado e é enviado pelo invasor para um servidor intermediário. O servidor é levado a enviar seus pacotes de resposta UDP para o IP da vítima de destino, em vez de voltar para o endereço IP do invasor. O servidor intermediário é usado porque gera uma resposta várias vezes maior do que o pacote de solicitação, amplificando efetivamente a quantidade de tráfego de ataque enviado ao endereço IP de destino.

O fator de amplificação é a razão entre o tamanho da resposta e o tamanho da solicitação e varia dependendo do protocolo usado pelo invasor: DNS, NTP, SSDP, CLDAP, Memcached, CharGen ou QOTD. Por exemplo, o fator de amplificação para DNS pode ser de 28 a 54 vezes o número original de bytes. Portanto, se um invasor enviar uma carga útil de solicitação de 64 bytes para

um servidor DNS, ele poderá gerar mais de 3.400 bytes de tráfego indesejado para um destino de ataque. Os ataques de reflexão UDP são responsáveis por um maior volume de tráfego em comparação com outros ataques. A figura Ataque de reflexão UDP ilustra a tática de reflexão e o efeito de amplificação.

Ataque de reflexão de UDP

Ataques de inundação SYN

Quando um usuário se conecta a um serviço TCP (Transmission Control Protocol), como um servidor Web, seu cliente envia um pacote de sincronização SYN. O servidor retorna um pacote SYN-ACK em reconhecimento e, finalmente, o cliente responde com um pacote de confirmação (ACK), que completa o handshake de três vias esperado. A imagem a seguir ilustra esse handshake típico.

Handshake SYN em 3 vias

Em um ataque de inundação SYN, um cliente mal-intencionado envia um grande número de pacotes SYN, mas nunca envia os pacotes ACK finais para completar os handshakes. O servidor fica aguardando uma resposta às conexões TCP semiabertas e, eventualmente, fica sem capacidade para aceitar novas conexões TCP. Isso pode impedir que novos usuários se conectem ao servidor. O ataque está tentando amarrar as conexões de servidor disponíveis para que os recursos não estejam disponíveis para conexões legítimas. Embora as inundações de SYN possam atingir até centenas de Gbps, o objetivo do ataque não é aumentar o volume de tráfego SYN.

Ataques da camada de aplicação

Um invasor pode direcionar a própria aplicação usando um ataque de camada 7 ou camada de aplicação. Nesses ataques, semelhantes aos ataques de infraestrutura de inundação SYN, o invasor tenta sobrecarregar funções específicas de uma aplicação para tornar a aplicação indisponível ou não responde a usuários legítimos. Às vezes, isso pode ser alcançado com volumes de solicitações muito baixos que geram apenas um pequeno volume de tráfego de rede. Isso pode tornar o ataque difícil de detectar e mitigar. Exemplos de ataques de camada de aplicação incluem inundações HTTP, ataques de quebra de cache e inundações WordPress XML-RPC.

Em um ataque de inundação HTTP, um invasor envia solicitações HTTP que parecem ser de um usuário válido da aplicação Web. Algumas inundações HTTP têm como alvo um recurso específico,

enquanto inundações HTTP mais complexas tentam emular a interação humana com a aplicação. Isso pode aumentar a dificuldade de usar técnicas comuns de mitigação, como limitação da taxa de solicitação.

Os ataques de quebra de cache são um tipo de inundação HTTP que usa variações na string de consulta para contornar o armazenamento em cache da rede de entrega de conteúdo (CDN). Em vez de poder retornar resultados armazenados em cache, o CDN deve entrar em contato com o servidor de origem para cada solicitação de página, e esses downloads da origem causam tensão adicional no servidor da web da aplicação.

Com um ataque de inundação WordPress XML-RPC, também conhecido como inundação de pingback do WordPress, um invasor tem como destino um site hospedado no software de gerenciamento de conteúdo do WordPress. O invasor usa indevidamente a função da API XML-RPC para gerar uma inundação de solicitações HTTP. O recurso de pingback permite que um site hospedado no WordPress (Site A) notifique um site do WordPress diferente (Site B) por meio de um link que o Site A criou para o Site B. O Site B tenta buscar o Site A para verificar a existência do link. Em uma inundação de pingback, o invasor usa indevidamente esse recurso para fazer com que o Site B ataque o Site A. Esse tipo de ataque tem uma assinatura clara: o WordPress geralmente está presente no User-Agent do cabeçalho de solicitação HTTP.

Existem outras formas de tráfego malicioso que podem afetar a disponibilidade de uma aplicação. Os bots do Scraper automatizam as tentativas de acessar uma aplicação Web para roubar conteúdo ou registrar informações competitivas, como preços. Ataques de força bruta e preenchimento de credenciais são esforços programados para obter acesso não autorizado a áreas seguras de uma aplicação. Esses ataques não são estritamente de DDoS; mas sua natureza automatizada pode ser semelhante a um ataque de DDoS e eles podem ser mitigados implementando algumas das mesmas práticas recomendadas a serem abordadas neste documento.

Os ataques da camada de aplicação também podem ter como alvo serviços de Sistema de Nomes de Domínio (DNS). O mais comum desses ataques é uma inundação de consultas DNS em que um invasor usa muitas consultas de DNS bem formadas para esgotar os recursos de um servidor DNS. Esses ataques também podem incluir um componente de eliminação de cache em que o invasor randomiza a string de subdomínio para ignorar o cache DNS local de qualquer resolvedor. Como resultado, o resolvedor não pode tirar proveito das consultas de domínio em cache e deve entrar em contato repetidamente com o servidor DNS autoritativo, o que amplifica o ataque.

Se uma aplicação Web for entregue por meio do Transport Layer Security (TLS), um invasor também poderá optar por atacar o processo de negociação de TLS. O TLS é computacionalmente caro, portanto, um invasor, ao gerar workload extra no servidor para processar dados ilegíveis

(ou ininteligíveis (texto cifrado)) como um handshake legítimo, pode reduzir a disponibilidade do servidor. Em uma variação desse ataque, um invasor conclui o handshake TLS, mas renegocia perpetuamente o método de criptografia. Como alternativa, um invasor pode tentar esgotar os recursos do servidor abrindo e fechando muitas sessões TLS.

Técnicas de mitigação

Algumas formas de mitigação de DDoS são incluídas automaticamente nos serviços da AWS. A resiliência de DDoS pode ser aprimorada ainda mais usando uma arquitetura da AWS com serviços específicos, abordados nas seções a seguir, e implementando práticas recomendadas adicionais para cada parte do fluxo de rede entre os usuários e sua aplicação.

Todos os clientes da AWS podem se beneficiar das proteções automáticas do AWS Shield Standard sem custos adicionais. O AWS Shield Standard protege contra os ataques de DDoS mais comuns e frequentes, que ocorrem na rede e na camada de transporte, e afetam seus sites ou aplicações. Essa proteção está sempre ativa, pré-configurada, estática e não fornece relatórios ou análises. Ela é oferecida em todos os serviços da AWS e em todas as regiões da AWS. Nas regiões da AWS, os ataques de DDoS são detectados e o sistema Shield Standard define automaticamente o tráfego de referência, identifica anomalias e, conforme necessário, cria mitigações. Você pode usar o AWS Shield Standard como parte de uma arquitetura resistente a DDoS para proteger aplicações Web e não Web.

Você também pode utilizar os serviços da AWS que operam de locais da borda, como Amazon CloudFront, Global Accelerator e Route 53 para criar proteção de disponibilidade abrangente contra todos os ataques conhecidos da camada de infraestrutura. Esses serviços fazem parte da AWS Global Edge Network e podem melhorar a resiliência de DDoS da sua aplicação ao servir qualquer tipo de tráfego de aplicações de pontos de presença distribuídos em todo o mundo. Você pode executar sua aplicação em qualquer região da AWS e usar esses serviços para proteger a disponibilidade da aplicação e otimizar a performance da aplicação para usuários finais legítimos.

Os benefícios do uso do Amazon CloudFront, Global Accelerator e Amazon Route 53 incluem:

- Acesso à Internet e à capacidade de mitigação de DDoS em toda a AWS Global Edge Network. Isso é útil para mitigar ataques volumétricos maiores, que podem atingir a escala de terabits.
- Os sistemas de mitigação de DDoS do AWS Shield são integrados aos serviços de borda da AWS, reduzindo o tempo de mitigação de minutos para menos de um segundo.
- As técnicas de mitigação de inundação SYN sem estado fazem proxy e verificam as conexões de entrada antes de passá-las para o serviço protegido. Isso garante que apenas conexões válidas cheguem à sua aplicação e, ao mesmo tempo, protejam seus usuários finais legítimos contra descartes de falsos positivos.

- Sistemas automáticos de engenharia de tráfego que dispersam ou isolam o impacto de grandes ataques de DDoS volumétricos. Todos esses serviços isolam os ataques na origem antes que eles atinjam sua origem, o que significa menos impacto nos sistemas protegidos por esses serviços.
- A defesa da camada de aplicação, quando combinada com o AWS WAF, não requer alteração da arquitetura atual da aplicação (por exemplo, em uma região da AWS ou em um datacenter on-premises).

Não há cobrança pela transferência de dados de entrada na AWS e você não paga pelo tráfego de ataque de DDoS que é mitigado pelo AWS Shield. O diagrama de arquitetura a seguir inclui os serviços da AWS Global Edge Network.

Essa arquitetura inclui vários serviços da AWS que podem ajudá-lo a melhorar a resiliência do sua aplicação Web contra ataques de DDoS. A tabela Resumo das práticas recomendadas fornece um resumo desses serviços e dos recursos que eles podem oferecer. A AWS marcou cada serviço com um indicador de práticas recomendadas (BP1, BP2) para facilitar a referência neste documento. Por exemplo, uma próxima seção discutirá os recursos fornecidos pelo Amazon CloudFront e pelo Global Accelerator que inclui o indicador BP1 de práticas recomendadas.

Tabela 2: resumo das práticas recomendadas

AWS Edge	Região da AWS					
	Uso do Amazon CloudFront (BP1) com o AWS WAF (BP2)	Uso do Global Accelerator (BP1)	Como usar o Amazon Route 53 (BP3)	Uso do Elastic Load Balancing (BP6) com o AWS WAF (BP2)	Uso de grupos de segurança e ACLs de rede na Amazon VPC (BP5)	Uso do Amazon EC2 Auto Scaling (BP7)
Mitigação de ataques da camada 3 (por exemplo,	✓	✓	✓	✓	✓	✓

AWS Edge	Região da AWS					
reflexão UDP)						
Mitigação de ataques da camada 4 (por exemplo, inundação SYN)	✓	✓	✓	✓		
Mitigação de ataques da camada 6 (por exemplo, TLS)	✓	✓	✓	✓		
Reduzir superfície de ataque	✓	✓	✓	✓	✓	
Escale para absorver o tráfego da camada de aplicação	✓	✓	✓	✓	✓	✓
Mitigação de ataques da camada 7 (camada de aplicação)	✓	✓(*)	✓	✓	✓(*)	✓(*)

AWS Edge	Região da AWS					
Isolamento geográfico e dispersão do excesso de tráfego e ataques de DDoS maiores	✓	✓	✓			
✓(*): se usado com o AWS WAF com o Application Load Balancer						

Outra maneira de melhorar sua prontidão para responder e mitigar ataques de DDoS é a inscrição no AWS Shield Advanced.

Os clientes recebem detecção personalizada com base em:

- Padrões de tráfego específicos da sua aplicação.
- Proteção contra ataques de DDoS de camada 7, inclusive o AWS WAF sem custo adicional.
- Acesso ao suporte especializado 24x7 do AWS SRT.
- Gerenciamento centralizado de políticas de segurança por meio do AWS Firewall Manager.
- Proteção de custos para resguardo contra cobranças de dimensionamento resultantes de picos de uso relacionados a DDoS.

Esse serviço opcional de mitigação de DDoS ajuda a proteger aplicações hospedadas em qualquer região da AWS. O serviço está disponível globalmente para CloudFront, Route 53 e Global

Accelerator. O uso do Shield Advanced com endereços de IP elástico permite proteger instâncias do Network Load Balancer (NLBs) ou do Amazon EC2.

Os benefícios do uso do AWS Shield Advanced incluem:

- Acesso ao AWS SRT para obter assistência na mitigação de ataques de DDoS que afetam a disponibilidade da aplicação.
- Visibilidade de ataques de DDoS usando as métricas e alarmes do AWS Management Console, da API e do Amazon CloudWatch.
- Acesso ao histórico de todos os eventos de DDoS dos últimos 13 meses.
- Acesso ao firewall de aplicações Web da AWS (AWS WAF), sem custo adicional para a mitigação de ataques de DDoS na camada de aplicação (quando usado com o Amazon CloudFront ou o Application Load Balancer).
- Linha de base automática de atributos de tráfego da Web, quando usado com o AWS WAF.
- Acesso ao AWS Firewall Manager, sem custo adicional, para imposição automatizada de políticas.
- Limites de detecção restritos que encaminham o tráfego para o sistema de mitigação de DDoS mais cedo e podem melhorar o tempo de mitigação de ataques contra o Amazon EC2 ou o Network Load Balancer, quando usados com um endereço de IP elástico.
- Proteção de custos que permite solicitar um reembolso limitado dos custos relacionados à escalabilidade resultantes de um ataque de DDoS.
- Acordo de nível de serviço aprimorado específico de clientes do AWS Shield Advanced.
- Engajamento proativo do AWS SRT quando um evento Shield é detectado.
- Grupos de proteção que permitem empacotar recursos, fornecendo uma forma de autoatendimento para personalizar o escopo de detecção e mitigação de sua aplicação, tratando vários recursos como uma única unidade. O agrupamento de recursos melhora a precisão da detecção, minimiza falsos positivos, facilita a proteção automática de recursos recém-criados e acelera o tempo para mitigar ataques contra muitos recursos que compõem uma única aplicação. Para obter informações sobre grupos de proteção, consulte Grupos de proteção Shield Advanced.

Para obter uma lista completa de recursos do AWS Shield Advanced e obter mais informações sobre o AWS Shield, consulte Funcionamento do AWS Shield.

Tópicos

- Práticas recomendadas para mitigação de DDoS
- Aproveite os pontos de presença da AWS para escala (BP1, BP3)

- Defesa da camada de aplicação (BP1, BP2)

Práticas recomendadas para mitigação de DDoS

Nas seções a seguir, cada uma das práticas recomendadas para mitigação de DDoS é descrita em mais detalhes. Para obter um guia rápido e fácil de implementar sobre como criar uma camada de mitigação de DDoS para aplicações Web estáticas ou dinâmicas, consulte How to Help Protect Dynamic Web Applications Against DDoS Attacks (Como ajudar a proteger aplicações Web dinâmicas contra ataques de DDoS).

Defesa da camada de infraestrutura (BP1, BP3, BP6, BP7)

Em um ambiente de datacenter tradicional, você pode mitigar ataques de DDoS na camada de infraestrutura usando técnicas como o provisionamento de capacidade em excesso, implantação de sistemas de mitigação de DDoS ou depuração de tráfego com a ajuda de serviços de mitigação de DDoS. Na AWS, os recursos de mitigação de DDoS são fornecidos automaticamente; mas você pode otimizar a resiliência de DDoS da sua aplicação fazendo escolhas de arquitetura que aproveitam melhor esses recursos e também permitem que você escale para o excesso de tráfego.

As principais considerações para ajudar a mitigar ataques de DDoS volumétricos incluem garantir que capacidade de trânsito e diversidade suficientes estejam disponíveis e proteger recursos da AWS, como instâncias do Amazon EC2, contra o tráfego de ataque.

Alguns tipos de instância do Amazon EC2 oferecem suporte a recursos que podem lidar mais facilmente com grandes volumes de tráfego, por exemplo, interfaces de largura de banda de rede de até 100 Gbps e redes aprimoradas. Isso ajuda a evitar o congestionamento da interface para o tráfego que atingiu a instância do Amazon EC2. As instâncias que oferecem suporte a redes aprimoradas oferecem maior performance de E/S, maior largura de banda e menor utilização da CPU em comparação com as implementações tradicionais. Isso melhora a capacidade da instância de lidar com grandes volumes de tráfego e, por fim, os torna altamente resilientes contra a carga de pacotes por segundo (pps).

Para permitir esse alto nível de resiliência, a AWS recomenda o uso de instâncias dedicadas do Amazon EC2 ou instâncias do Amazon EC2 com taxa de transferência de rede mais alta que tenham um sufixo N e suporte para redes aprimoradas com até 100 Gbps de largura de banda de rede, por exemplo, c6gn.16xlarge e c5n.18xlarge ou metal instâncias (como c5n.metal).

Para obter mais informações sobre instâncias do Amazon EC2 que oferecem suporte a interfaces de rede de 100 Gigabit e redes aprimoradas, consulte Tipos de instância do Amazon EC2.

O módulo necessário para redes avançadas e o conjunto de atributos enaSupport necessários estão incluídos no Amazon Linux 2 e nas versões mais recentes do Amazon Linux AMI. Portanto, se você iniciar uma instância com uma versão HVM do Amazon Linux em um tipo de instância compatível, a rede avançada já estará habilitada para a instância. Para obter mais informações, consulte Testar se a rede aprimorada está habilitada. Para obter mais informações sobre como habilitar redes aprimoradas, consulte Rede aprimorada no Linux.

Amazon EC2 com Auto Scaling (BP7)

Outra maneira de mitigar os ataques à infraestrutura e à camada de aplicação é operar em escala. Se você tiver aplicações Web, poderá usar balanceadores de carga para distribuir o tráfego para várias instâncias do Amazon EC2 que são provisionadas em excesso ou configuradas para escalar automaticamente. Essas instâncias podem lidar com picos repentinos de tráfego que ocorrem por qualquer motivo, incluindo uma multidão de flash ou um ataque de DDoS na camada de aplicação. Você pode definir alarmes do Amazon CloudWatch para iniciar o Auto Scaling para escalar automaticamente o tamanho da sua frota do Amazon EC2 em resposta a eventos que você define, como CPU, RAM, E/S de rede e até mesmo métricas personalizadas. Essa abordagem protege a disponibilidade das aplicações quando há um aumento inesperado no volume de solicitações. Ao usar o Amazon CloudFront, Application Load Balancer, Classic Load Balancers ou Network Load Balancer com sua aplicação, a negociação de TLS é feita pela distribuição (Amazon CloudFront) ou pelo balanceador de carga. Esses recursos ajudam a proteger suas instâncias de serem afetadas por ataques baseados em TLS, escalonando para lidar com solicitações legítimas e ataques de uso abusivo de TLS.

Para obter mais informações sobre como usar o Amazon CloudWatch para invocar o Auto Scaling, consulte Monitoramento de métricas do Amazon CloudWatch para seus grupos e instâncias do Auto Scaling.

O Amazon EC2 fornece capacidade computacional redimensionável para que você possa aumentar ou diminuir a escala verticalmente com rapidez conforme os requisitos mudam. Você pode escalar horizontalmente adicionando instâncias automaticamente à sua aplicação ao Escalabilidade do tamanho do seu grupo do Amazon EC2 Auto Scaling, e pode escalar verticalmente usando tipos de instância do EC2 maiores.

Elastic Load Balancing (BP6)

Grandes ataques de DDoS podem sobrecarregar a capacidade de uma única instância do Amazon EC2. Com o Elastic Load Balancing (ELB), você pode reduzir o risco de sobrecarregar sua aplicação

distribuindo o tráfego entre várias instâncias de backend. O Elastic Load Balancing pode ser escalado automaticamente, permitindo que você gerencie volumes maiores quando houver tráfego extra imprevisto, por exemplo, devido a flash crowds ou ataques de DDoS. Para aplicações criadas em uma Amazon VPC, há três tipos de ELBs a serem considerados, dependendo do tipo de aplicação: Application Load Balancer (ALB), Classic Load Balancer (CLB) e Network Load Balancer (NLB).

Para aplicações Web, você pode usar o Application Load Balancer para rotear o tráfego com base no conteúdo e aceitar somente solicitações da Web bem formadas. O Application Load Balancer bloqueia muitos ataques de DDoS comuns, como inundações SYN ou ataques de reflexão UDP, protegendo sua aplicação do ataque. O Application Load Balancer é escalado automaticamente para absorver o tráfego adicional quando esses tipos de ataques são detectados. As ações de escalabilidade devido a ataques na camada de infraestrutura são transparentes para os clientes da AWS e não afetam sua fatura.

Para obter mais informações sobre a proteção de aplicações Web com o Application Load Balancer, consulte Conceitos básicos dos Application Load Balancers

Para aplicações baseadas em TCP, você pode usar o Network Load Balancer para rotear o tráfego para destinos (por exemplo, instâncias do Amazon EC2) com latência ultrabaixa. Uma consideração importante com o Network Load Balancer é que qualquer tráfego que chegar ao balanceador de carga em um ouvinte válido será roteado para seus destinos, não absorvido. Você pode usar o Shield Advanced para configurar a proteção contra DDoS para endereços de IP elástico. Quando um endereço de IP elástico é atribuído por zona de disponibilidade ao Network Load Balancer, o Shield Advanced aplicará as proteções DDoS relevantes para o tráfego do Network Load Balancer.

Para obter mais informações sobre a proteção de aplicações TCP com o Network Load Balancer, consulte Introdução aos Network Load Balancers

Aproveite os pontos de presença da AWS para escala (BP1, BP3)

O acesso a conexões de Internet diversificadas e altamente escaladas pode aumentar significativamente sua capacidade de otimizar a latência e a taxa de transferência para os usuários, absorver ataques de DDoS e isolar falhas, minimizando o impacto na disponibilidade da sua aplicação. Os locais da borda da AWS fornecem uma camada adicional de infraestrutura de rede que fornece esses benefícios a qualquer aplicação Web que usa o Amazon CloudFront, o Global Accelerator e o Amazon Route 53. Com esses serviços, você pode proteger de forma abrangente na borda suas aplicações executadas nas regiões da AWS.

Entrega de aplicação Web na borda (BP1)

O Amazon CloudFront é um serviço que pode ser usado para entregar todo o seu site, incluindo conteúdo estático, dinâmico, de transmissão e interativo. Conexões persistentes e configurações de tempo de vida variável (TTL) podem ser usadas para descarregar o tráfego de sua origem, mesmo que você não esteja veiculando conteúdo armazenável em cache. O uso desses recursos do CloudFront reduz o número de solicitações e conexões TCP de volta à sua origem, ajudando a proteger sua aplicação Web contra inundações HTTP. O CloudFront só aceita conexões bem formadas, o que ajuda a impedir que muitos ataques de DDoS comuns, como inundações SYN e ataques de reflexão UDP, atinjam sua origem. Os ataques de DDoS também são isolados geograficamente próximos à origem, o que evita que o tráfego afete outros locais. Todos esses recursos podem melhorar muito a capacidade de continuar oferecendo tráfego a usuários durante grandes ataques de DDoS. Você pode usar o CloudFront para proteger uma origem na AWS ou em outro lugar da Internet.

Se você estiver usando o Amazon S3 para servir conteúdo estático na Internet, a AWS recomenda usar o Amazon CloudFront para proteger seu bucket. Você pode usar a identificação de acesso de origem (OAI) para garantir que os usuários acessem apenas seus objetos usando URLs do CloudFront.

Para obter mais informações sobre o OAI, consulte Restringir o acesso ao conteúdo do Amazon S3 usando uma identidade de acesso de origem.

Para obter mais informações sobre como proteger e otimizar a performance de aplicações Web com o Amazon CloudFront, consulte Conceitos básicos do CloudFront.

Proteja ainda mais o tráfego de rede da sua origem usando o AWS Global Accelerator (BP1)

O Global Accelerator é um serviço de rede que melhora a disponibilidade e a performance do tráfego dos usuários em até 60%. Isso é feito deixando o tráfego entrar no ponto de presença mais próximo dos usuários e roteando-o pela infraestrutura de rede global da AWS para sua aplicação, seja ele executado em uma única ou em várias regiões da AWS.

O Global Accelerator encaminha o tráfego TCP e UDP para o endpoint ideal com base na performance na região da AWS mais próxima do usuário. Se houver uma falha na aplicação, o Global Accelerator fornecerá failover para o próximo melhor endpoint em 30 segundos. O Global Accelerator usa a vasta capacidade da rede global da AWS e as integrações com o Shield, como

um recurso de proxy SYN sem estado que desafia novas tentativas de conexão e atende apenas usuários finais legítimos, para proteger aplicações.

Você pode implementar uma arquitetura resiliente de DDoS que ofereça muitos dos mesmos benefícios que o Web Application Delivery nas práticas recomendadas de Edge, mesmo que sua aplicação use protocolos sem suporte do CloudFront ou se você estiver operando uma aplicação Web que exija endereços IP estáticos globais. Por exemplo, você pode exigir endereços IP que seus usuários finais possam adicionar à lista de permissões em seus firewalls e não sejam usados por nenhum outro cliente da AWS. Nesses cenários, você pode usar o Global Accelerator para proteger aplicações Web em execução no Application Load Balancer e em conjunto com o AWS WAF para também detectar e mitigar inundações de solicitação da camada de aplicação Web.

Para obter mais informações sobre como proteger e otimizar a performance do tráfego de rede usando o Global Accelerator, consulte Primeiros passos com o Global Accelerator.

Resolução de nome de domínio na borda (BP3)

O Amazon Route 53 é um serviço de Sistema de Nomes de Domínio (DNS) altamente disponível e escalável que pode ser usado para direcionar o tráfego para sua aplicação Web. Ele inclui recursos avançados como fluxo de tráfego, verificações e monitoramento de integridade, encaminhamento por latência e Geo DNS. Esses recursos avançados permitem controlar como o serviço responde às solicitações de DNS para melhorar a performance da sua aplicação Web e evitar interrupções no site.

O Amazon Route 53 usa técnicas como fragmentação aleatória e striping anycast, que podem ajudar os usuários a acessar sua aplicação mesmo se o serviço DNS for alvo de um ataque de DDoS.

Com a fragmentação aleatória, cada servidor de nomes em seu conjunto de delegação corresponde a um conjunto exclusivo de pontos de presença e caminhos da Internet. Isso proporciona maior tolerância a falhas e minimiza a sobreposição entre os clientes. Se um servidor de nomes no conjunto de delegações não estiver disponível, os usuários poderão tentar novamente e receber uma resposta de outro servidor de nomes em um local da borda diferente.

O striping Anycast permite que cada solicitação de DNS seja atendida pelo local mais ideal, dispersando a carga da rede e reduzindo a latência do DNS. Isso fornece uma resposta mais rápida para os usuários. Além disso, o Amazon Route 53 pode detectar anomalias na origem e no volume de consultas de DNS e priorizar solicitações de usuários que são conhecidos como confiáveis.

Para obter mais informações sobre como usar o Amazon Route 53 para rotear usuários para sua aplicação, consulte Conceitos básicos do Amazon Route 53.

Defesa da camada de aplicação (BP1, BP2)

Muitas das técnicas discutidas até agora neste documento são eficazes para mitigar o impacto que os ataques de DDoS da camada de infraestrutura têm na disponibilidade da sua aplicação. Para também se defender contra ataques na camada de aplicação, você precisa implementar uma arquitetura que permita detectar, dimensionar especificamente para absorver e bloquear solicitações mal-intencionadas. Essa é uma consideração importante porque os sistemas de mitigação de DDoS baseados em rede geralmente são ineficazes na mitigação de ataques complexos da camada de aplicação.

Detectar e filtrar solicitações da web mal-intencionadas (BP1, BP2)

Quando sua aplicação for executada na AWS, você poderá aproveitar o Amazon CloudFront e o AWS WAF para ajudar a se defender contra ataques de DDoS na camada de aplicação.

O Amazon CloudFront permite que você armazene em cache conteúdo estático e o disponibilize a partir de locais da borda da AWS, o que pode ajudar a reduzir a carga na sua origem. Também pode ajudar a reduzir a carga do servidor, impedindo que o tráfego que não seja da Web chegue à sua origem. Além disso, o CloudFront pode fechar automaticamente conexões de invasores de leitura lenta ou gravação lenta (por exemplo, Slowloris).

Usando o AWS WAF, você pode configurar listas de controle de acesso à Web (ACLs da Web) em suas distribuições do CloudFront ou Application Load Balancers para filtrar e bloquear solicitações com base em assinaturas de solicitação. Cada ACL da Web consiste em regras que você pode configurar para corresponder a string ou regex corresponder a um ou mais atributos de solicitação, como URI (Uniform Resource Identifier), string de consulta, método HTTP ou chave de cabeçalho. Além disso, usando as regras baseadas em taxa do AWS WAF, você pode bloquear automaticamente os endereços IP de agentes mal-intencionados quando as solicitações correspondentes a uma regra excederem um limite definido por você.

Solicitações de endereços IP de clientes infratores receberão as respostas do erro 403 Proibido e permanecerão bloqueadas até que as taxas de solicitação fiquem abaixo do limite. Isso é útil para mitigar ataques de inundação HTTP disfarçados como tráfego regular na Web. Para bloquear ataques com base na reputação do endereço IP, você pode criar regras usando condições de correspondência de IP ou usar as Regras Gerenciadas para o AWS WAF oferecidas pelos vendedores no AWS Marketplace. O AWS WAF oferece diretamente o AWS Managed Rules como um serviço gerenciado no qual você pode escolher grupos de regras de reputação de IP. O grupo de regras da lista de reputação de IP da Amazon contém regras baseadas na inteligência de ameaças

internas da Amazon. Isso será útil se você quiser bloquear endereços IP normalmente associados a bots ou outras ameaças. Esse grupo de regras da lista de IPs anônimos contém regras para bloquear solicitações de serviços que permitem a ofuscação da identidade do visualizador. Isso inclui solicitações de VPNs, proxies, nós Tor e plataformas de nuvem (incluindo a AWS). O AWS WAF e o CloudFront também permitem que você defina restrições geográficas para bloquear ou permitir solicitações de países selecionados. Isso pode ajudar a bloquear ataques de localizações geográficas onde você não espera atender usuários.

Para ajudar a identificar solicitações mal-intencionadas, revise os logs do servidor da Web ou use os recursos de registro e Solicitações de Amostra do AWS WAF. Ao habilitar os logs do AWS WAF, você obtém informações detalhadas sobre o tráfego analisado pela Web ACL. O AWS WAF oferece suporte à filtragem de log, permitindo que você especifique quais solicitações da web são registradas e quais solicitações são descartadas do log após a inspeção.

As informações registradas nos logs incluem a hora em que o AWS WAF recebeu a solicitação do seu recurso da AWS recurso, informações detalhadas sobre a solicitação e a ação correspondente para cada regra solicitada. Solicitações de amostra fornecem detalhes sobre solicitações nas últimas três horas que correspondem a uma de suas regras do AWS WAF. Você pode usar essas informações para identificar assinaturas de tráfego potencialmente mal-intencionadas e criar uma nova regra para negar essas solicitações. Se você vir várias solicitações com uma string de consulta aleatória, permita apenas os parâmetros da string de consulta relevantes para o cache da sua aplicação. Essa técnica é útil para mitigar um ataque de quebra de cache contra sua origem.

Se você estiver inscrito no AWS Shield Advanced, poderá contratar o AWS Shield Response Team (SRT) para ajudá-lo a criar regras para mitigar um ataque que esteja prejudicando a disponibilidade da sua aplicação. Você pode conceder acesso limitado ao AWS SRT ao Shield Advanced da sua conta e às APIs do AWS WAF. O AWS SRT acessa essas APIs para colocar mitigações em sua conta somente com sua autorização explícita. Para obter mais informações, consulte a seção Suporte deste documento.

Você pode usar o AWS Firewall Manager para configurar e gerenciar centralmente regras de segurança, como proteções do Shield Advanced e regras do AWS WAF, em toda a sua organização. Sua conta de gerenciamento do AWS Organizations pode designar uma conta de administrador, que está autorizada a criar políticas do Firewall Manager. Essas políticas permitem definir critérios, como tipo de recurso e etiquetas, que determinam onde as regras são aplicadas. Isso é útil quando você tem várias contas e deseja padronizar sua proteção.

Para obter mais informações sobre:

- Regras gerenciadas da AWS para o AWS WAF, consulte Regras gerenciadas da AWS para o AWS WAF.
- Usando a restrição geográfica para limitar o acesso à sua distribuição do CloudFront, consulte Restringir a distribuição geográfica do seu conteúdo.
- Uso do AWS WAF, consulte
  - Conceitos básicos do AWS WAF
  - Registrar em log as informações de tráfego da web ACL
  - Visualizar um exemplo de solicitações da web
- Configuração de regras baseadas em taxa, consulte Proteger sites e serviços usando regras baseadas em taxa para o AWS WAF
- Como gerenciar a implantação de regras do AWS WAF em seus recursos da AWS com o Firewall Manager, consulte
  - Primeiros passos com as políticas do AWS WAF do Firewall Manager.
  - Primeiros passos com as políticas Shield Advanced do Firewall Manager.

Redução da superfície de ataque

Outra consideração importante ao arquitetar uma solução da AWS é limitar as oportunidades que um invasor tem para escolher sua aplicação como destino. Esse conceito é conhecido como redução da superfície de ataque. Os recursos que não estão expostos à Internet são mais difíceis de atacar, o que limita as opções que um invasor tem para atingir a disponibilidade da sua aplicação.

Por exemplo, se você não espera que os usuários interajam diretamente com determinados recursos, certifique-se de que esses recursos não estejam acessíveis pela Internet. Da mesma forma, não aceite tráfego de usuários ou aplicações externas em portas ou protocolos que não sejam necessários para comunicação.

Na seção a seguir, a AWS fornece práticas recomendadas para orientar você na redução da superfície de ataque e na limitação da exposição da aplicação à Internet.

Tópicos

Ofuscar recursos do AWS (BP1, BP4, BP5)

Normalmente, os usuários podem usar uma aplicação de forma rápida e fácil sem exigir que os recursos da AWS sejam totalmente expostos à Internet. Por exemplo, quando você tem instâncias do Amazon EC2 por trás de um Elastic Load Balancing, as instâncias em si podem não precisar ser acessíveis publicamente. Em vez disso, você pode fornecer aos usuários acesso ao Elastic Load Balancing em determinadas portas TCP e permitir que apenas o Elastic Load Balancing se comunique com as instâncias. Você pode fazer isso ao configurar grupos de segurança e listas de controle de acesso de rede (NACLs) dentro da Amazon Virtual Private Cloud (Amazon VPC). A Amazon VPC permite provisionar uma seção isolada logicamente da Nuvem AWS, em que você pode executar recursos da AWS em uma rede virtual definida.

Os grupos de segurança e as ACLs de rede são semelhantes ao permitir controlar o acesso a recursos da AWS dentro da sua VPC. Mas os grupos de segurança permitem controlar o tráfego de entrada e saída no nível da instância, enquanto as ACLs de rede oferecem recursos semelhantes no nível da sub-rede da VPC. Não há cobrança adicional pelo uso de grupos de segurança ou ACLs de rede.

Grupos de segurança e listas de controle de acesso de rede (ACLs de rede) (BP5)

Você pode escolher se deseja especificar grupos de segurança ao executar uma instância ou associar a instância a um grupo de segurança posteriormente. Todo o tráfego para um grupo de segurança da Internet é negado implicitamente, a menos que você crie uma regra de permissão para permitir o tráfego. Por exemplo, se você tiver uma aplicação Web que usa um Elastic Load Balancing e várias instâncias do Amazon EC2, poderá optar por criar um grupo de segurança para o Elastic Load Balancing (grupo de segurança do Elastic Load Balancing) e um para as instâncias (grupo de segurança do servidor de aplicações Web). Em seguida, você poderá criar regras de permissão para permitir o tráfego da Internet para o grupo de segurança do ELB e o tráfego do grupo de segurança do ELB para o grupo de segurança do servidor da aplicação Web. Isso garante que o tráfego da Internet não poderá se comunicar diretamente com suas instâncias do Amazon EC2, o que dificulta para um invasor aprender como afetar sua aplicação.

Ao criar ACLs de rede, você poderá especificar as regras de permissão e de negação. Isso será útil caso você queira negar explicitamente determinados tipos de tráfego para sua aplicação. Por exemplo, você pode definir endereços IP (como intervalos CIDR), protocolos e portas de destino que devem ter acesso negado para toda a sub-rede. Se a aplicação for usada somente para o tráfego TCP, você poderá criar uma regra para negar todo o tráfego UDP, ou vice-versa. Essa opção é útil ao responder a ataques de DDoS porque permite criar suas próprias regras para mitigar o ataque quando você conhece os IPs de origem ou outra assinatura.

Se você estiver inscrito no AWS Shield Advanced, poderá registrar endereços de IP elástico como recursos protegidos. Os ataques de DDoS contra endereços de IP elástico que foram registrados como recursos protegidos são detectados mais rapidamente, o que pode resultar em um tempo mais rápido de mitigação. Quando um ataque é detectado, os sistemas de mitigação de DDoS leem a ACL de rede que corresponde ao IP elástico de destino e a aplicam na borda da rede da AWS. Isso reduz significativamente o risco de impacto de vários ataques de DDoS na camada de infraestrutura.

Para obter mais informações sobre como configurar grupos de segurança e ACLs de rede para otimizar a resiliência de DDoS, consulte How to Help Prepare for DDoS Attacks by Reducing Your Attack Surface (Como ajudar a se preparar para ataques de DDoS reduzindo sua superfície de ataque).

Para obter mais informações sobre como usar o Shield Advanced com endereços de IP elástico como recursos protegidos, consulte as etapas para Inscrever-se no AWS Shield Advanced.

Proteção da origem (BP1, BP5)

Se você estiver usando o Amazon CloudFront com uma origem que está dentro da sua VPC, convém garantir que somente sua distribuição do CloudFront possa encaminhar solicitações para sua origem. Com os cabeçalhos de solicitação Edge-to-Origin, você pode adicionar ou substituir o valor dos cabeçalhos de solicitação existentes quando o CloudFront encaminha solicitações para sua origem. Você pode usar os cabeçalhos personalizados de origem, por exemplo, o cabeçalho X-Shared-Secret, para ajudar a validar se as solicitações feitas à sua origem foram enviadas do CloudFront.

Para obter mais informações sobre como proteger sua origem com cabeçalhos personalizados de origem, consulte Como adicionar cabeçalhos personalizados a solicitações de origem e Como restringir o acesso a Application Load Balancers.

Para obter orientações sobre a implementação de uma solução de exemplo para alternar automaticamente o valor dos cabeçalhos personalizados de origem para a restrição de acesso de origem, consulte How to enhance Amazon CloudFront origin security with AWS WAF and Secrets Manager (Como aprimorar a segurança de origem do Amazon CloudFront com o AWS WAF e o Secrets Manager).

Como alternativa, você pode usar uma função do AWS Lambda para atualizar automaticamente as regras do grupo de segurança para permitir somente o tráfego do CloudFront. Isso melhora a segurança de sua origem ajudando a garantir que usuários mal-intencionados não possam ignorar o CloudFront e a AWS WAF ao acessarem sua aplicação Web.

Para obter mais informações sobre como proteger sua origem atualizando automaticamente seus grupos de segurança, consulte o cabeçalho X-Shared-Secret e a publicação How to Automatically Update Your Security Groups for Amazon CloudFront and AWS WAF by Using AWS Lambda (Como atualizar automaticamente seus grupos de segurança para o Amazon CloudFront e o AWS WAF usando o AWS Lambda).

Proteção de endpoints de API (BP4)

Normalmente, quando você precisa expor uma API ao público, existe o risco de o frontend da API ser alvo de um ataque de DDoS. Para ajudar a reduzir o risco, você pode usar o Amazon API Gateway como uma porta de entrada para aplicações executadas no Amazon EC2, no AWS Lambda ou em outro lugar. Ao usar o Amazon API Gateway, você não precisará de seus próprios servidores para o frontend de API e poderá ofuscar outros componentes da sua aplicação. Ao dificultar a detecção dos componentes da sua aplicação, você poderá ajudar a evitar que esses recursos da AWS sejam alvo de um ataque de DDoS.

Ao usar o Amazon API Gateway, você poderá escolher entre dois tipos de endpoints de API. A primeira é a opção padrão: endpoints de API otimizados para borda que são acessados por meio de uma distribuição do Amazon CloudFront. No entanto, a distribuição é criada e gerenciada pelo API Gateway, para que você não tenha controle sobre ela. A segunda opção é usar um endpoint de API regional que é acessado da mesma região da AWS em que sua API REST está implantada. A AWS recomenda que você use o segundo tipo de endpoint e o associe à sua própria distribuição do Amazon CloudFront. Isso lhe dá controle sobre a distribuição do Amazon CloudFront e a capacidade de usar o AWS WAF para proteção da camada de aplicação. Esse modo fornece acesso à capacidade de mitigação de DDoS escalada em toda a rede de borda global da AWS.

Ao usar o Amazon CloudFront e o AWS WAF com o Amazon API Gateway, configure as seguintes opções:

- Configure o comportamento do cache para suas distribuições para encaminhar todos os cabeçalhos para o endpoint regional do API Gateway. Ao fazer isso, o CloudFront tratará o conteúdo como dinâmico e ignorará o armazenamento em cache do conteúdo.
- Proteja seu API Gateway contra acesso direto configurando a distribuição para incluir o `x-api-key` do cabeçalho personalizado de origem, configurando o valor da chave da API no API Gateway.
- Proteja o backend contra tráfego excessivo configurando limites de taxa padrão ou de intermitência para cada método em suas APIs REST.

Para obter mais informações sobre a criação de APIs com o Amazon API Gateway, consulte Primeiros passos no Amazon API Gateway.

Técnicas operacionais

As técnicas de mitigação neste documento ajudam você a arquitetar aplicações que são inerentemente resilientes contra ataques de DDoS. Em muitos casos, também será útil saber quando um ataque de DDoS é direcionando à sua aplicação para que você possa tomar medidas de mitigação. Esta seção discute as práticas recomendadas para obter visibilidade sobre comportamento anormal, alertas e automação, gerenciamento de proteção em escala e envolvimento da AWS para suporte adicional.

Tópicos

Visibilidade

Quando uma métrica operacional importante se desvia substancialmente do valor esperado, um invasor pode estar tentando direcionar a disponibilidade da sua aplicação. A familiaridade com o comportamento normal da sua aplicação significa que você pode agir mais rapidamente ao detectar uma anomalia. O Amazon CloudWatch pode ajudar monitorando aplicações que você executa na AWS. Por exemplo, você pode coletar e rastrear métricas, coletar e monitorar arquivos de log, definir alarmes e responder automaticamente a alterações nos seus recursos da AWS.

Se você seguir a arquitetura de referência resistente a DDoS ao arquitetar sua aplicação, ataques comuns à camada de infraestrutura serão bloqueados antes de chegar à sua aplicação. Se você estiver inscrito no AWS Shield Advanced, terá acesso a várias métricas do CloudWatch que poderão indicar que sua aplicação é um destino. Por exemplo, você pode configurar alarmes para notificá-lo quando houver um ataque de DDoS em andamento, para que você possa verificar a integridade da sua aplicação e decidir se deseja envolver o AWS SRT. Você pode configurar a métrica do `DDoSDetected` para informar se um ataque foi detectado. Se você quiser ser alertado com base no volume de ataques, você também pode usar as métricas `DDoSAttackBitsPerSecond`, `DDoSAttackPacketsPerSecond` ou `DDoSAttackRequestsPerSecond`. Você pode monitorar essas métricas integrando o CloudWatch às suas próprias ferramentas ou usando ferramentas fornecidas por terceiros, como Slack ou PagerDuty.

Um ataque à camada de aplicação pode elevar muitas métricas do Amazon CloudWatch. Se estiver usando o AWS WAF, você poderá usar o CloudWatch para monitorar e ativar alarmes sobre aumentos nas solicitações que você definiu no AWS WAF para serem permitidas, contadas ou bloqueadas. Isso permite que você receba uma notificação se o nível de tráfego exceder o que sua aplicação pode manipular. Você também pode usar as métricas do Amazon CloudFront, do Amazon Route 53, do Application Load Balancer, do Network Load Balancer, do Amazon EC2 e do Auto Scaling que são rastreadas no CloudWatch para detectar alterações que podem indicar um ataque de DDoS.

A tabela Métricas recomendadas do CloudWatch lista descrições das métricas do CloudWatch que são comumente usadas para detectar e reagir a ataques de DDoS.

Tabela 3: métricas recomendadas do Amazon CloudWatch

Tópico	Métrica	Descrição
AWS Shield Advanced	DDoSDetected	Indica um evento de DDoS para um nome do recurso da Amazon (ARN) específico.
AWS Shield Advanced	DDoSAttackBitsPerSecond	O número de bytes observados durante um evento de DDoS para um ARN específico. Esta métrica está disponível apenas para a camada 3/4 dos eventos de DDoS.
AWS Shield Advanced	DDoSAttackPacketsPerSecond	O número de pacotes observados durante um evento de DDoS para um ARN específico. Esta métrica está disponível apenas para a camada 3/4 dos eventos de DDoS.
AWS Shield Advanced	DDoSAttackRequestsPerSecond	O número de solicitações observadas durante um evento de DDoS para um

Tópico	Métrica	Descrição
		ARN específico. Esta métrica está disponível apenas para a camada 7 dos eventos de DDoS e só será relatada para os eventos mais significativos da camada 7.
AWS WAF	AllowedRequests	O número de solicitações da web permitidas.
AWS WAF	BlockedRequests	O número de solicitações da web bloqueadas.
AWS WAF	CountedRequests	O número de solicitações da web contadas.
AWS WAF	PassedRequests	O número de solicitações aprovadas. Isso é usado apenas para solicitações que passam por uma avaliação de grupo de regras sem corresponder a nenhuma das regras do grupo de regras.
Amazon CloudFront	Solicitações	O número de solicitações HTTP/S.
Amazon CloudFront	TotalErrorRate	A porcentagem de todas as solicitações para a qual o código de status HTTP é 4xx ou 5xx.
Amazon Route 53	HealthCheckStatus	O status do endpoint de verificação de integridade.

Tópico	Métrica	Descrição
Application Load Balancer	ActiveConnectionCount	O número total de conexões TCP simultâneas ativas de clientes com o balanceador de carga e do balanceador de carga com destinos.
Application Load Balancer	ConsumedLCUs	O número de unidades de Load Balancer Capacity (LCU – Capacidade de balancead or de carga) usadas pelo balanceador de carga.
Application Load Balancer	HTTPCode_ELB_4XX_Count HTTPCode_ELB_5XX_Count	O número de códigos de erro do cliente HTTP 4xx ou 5xx gerados pelo balanceador de carga.
Application Load Balancer	NewConnectionCount	O número total de novas conexões TCP estabelecidas de clientes com o balanceador de carga e do balanceador de carga com destinos.
Application Load Balancer	ProcessedBytes	O número total de bytes processados pelo balanceador de carga.
Application Load Balancer	RejectedConnectionCount	O número de conexões que foram rejeitadas porque o balanceador de carga atingiu o número máximo de conexões.
Application Load Balancer	RequestCount	O número de solicitações que foram processadas.

Tópico	Métrica	Descrição
Application Load Balancer	TargetConnectionErrorCount	O número de conexões que não foram estabelecidas com êxito entre o balanceador de carga e o destino.
Application Load Balancer	TargetResponseTime	O tempo decorrido, em segundos, depois que a solicitação deixa o balancead or de carga até o momento em que uma resposta é recebida do destino.
Application Load Balancer	UnHealthyHostCount	O número de destinos considerados não íntegros.
Network Load Balancer	ActiveFlowCount	O número total de fluxos (conexões) TCP simultâneos dos clientes para os destinos.
Network Load Balancer	ConsumedLCUs	O número de unidades de Load Balancer Capacity (LCU – Capacidade de balancead or de carga) usadas pelo balanceador de carga.
Network Load Balancer	NewFlowCount	O número total de novos fluxos (ou conexões) TCP estabelecidos dos clientes para os destinos no período.
Network Load Balancer	ProcessedBytes	O número total de bytes processados pelo balanceador de carga, incluindo cabeçalho s TCP/IP.

Tópico	Métrica	Descrição
Global Accelerator	NewFlowCount	O número total de novos fluxos (ou conexões) TCP e UDP estabelecidos dos clientes para os endpoints no período.
Global Accelerator	ProcessedBytesIn	O número total de bytes recebidos processados pelo acelerador, incluindo cabeçalhos TCP/IP.
Auto Scaling	GroupMaxSize	O tamanho máximo do grupo de Auto Scaling.
Amazon EC2	CPUUtilization	O percentual de unidades de processamento EC2 alocadas que estão sendo utilizadas.
Amazon EC2	NetworkIn	A quantidade de bytes recebidos em todas as interfaces de rede pela instância.

Para obter mais informações sobre como usar o Amazon CloudWatch para detectar ataques de DDoS em sua aplicação, consulte Conceitos básicos do Amazon CloudWatch.

Para explorar um exemplo de um painel criado usando algumas das métricas da tabela anterior, consulte Um sistema de monitoramento de lista de referência personalizada

A AWS inclui várias métricas e alarmes adicionais para notificá-lo sobre um ataque e para ajudá-lo a monitorar os recursos da sua aplicação. O console do AWS Shield ou a API fornecem um resumo de eventos por conta e detalhes sobre os ataques que foram detectados.

Além disso, o painel do ambiente de ameaças globais fornece informações resumidas sobre todos os ataques de DDoS detectados pela AWS. Essas informações podem ser úteis para entender melhor

as ameaças de DDoS em uma população maior de aplicações, além das tendências de ataque, e comparar com os ataques que você pode ter observado.

Se você estiver inscrito no AWS Shield Advanced, o painel de serviço exibirá métricas adicionais de detecção e mitigação e detalhes do tráfego de rede para eventos detectados em recursos protegidos. O AWS Shield avalia o tráfego para seu recurso protegido em várias dimensões. Quando uma anomalia é detectada, o AWS Shield cria um evento e relata a dimensão de tráfego em que a anomalia foi observada. Com uma mitigação colocada, isso protege seu recurso de receber tráfego excessivo e tráfego que corresponda a uma assinatura de evento de DDoS conhecida.

As métricas de detecção são baseadas em fluxos de rede ou logs do AWS WAF de exemplo quando uma ACL da Web está associada ao recurso protegido. As métricas de mitigação são baseadas no tráfego observado pelos sistemas de mitigação de DDoS do Shield. As métricas de mitigação são uma medida mais precisa do tráfego em seu recurso.

A métrica dos principais colaboradores da rede fornece informações sobre a origem do tráfego durante um evento detectado. Você pode visualizar os colaboradores de maior volume e classificar por aspectos como protocolo, porta de origem e sinalizadores TCP. A métrica dos principais colaboradores inclui métricas para todo o tráfego observado no recurso ao longo de várias dimensões. Ele fornece dimensões métricas adicionais que você pode usar para entender o tráfego de rede que é enviado ao seu recurso durante um evento.

O painel de serviço também inclui detalhes sobre as ações executadas automaticamente para mitigar ataques de DDoS. Essas informações facilitam a investigação de anomalias, a exploração das dimensões do tráfego e a compreensão das ações tomadas pelo Shield Advanced para proteger sua disponibilidade.

Outra ferramenta que pode ajudar você a obter visibilidade do tráfego direcionado à sua aplicação são os Logs de fluxo da VPC. Em uma rede tradicional, você pode usar logs de fluxo de rede para solucionar problemas de conectividade e segurança e para garantir que as regras de acesso à rede estejam funcionando conforme o esperado. Usando os logs de fluxo da VPC, você pode capturar informações sobre o tráfego IP que está indo e vindo de interfaces de rede em sua VPC.

Cada registro de log de fluxo inclui o seguinte: endereços IP de origem e destino, portas de origem e destino, protocolo e o número de pacotes e bytes transferidos durante a janela de captura. Você pode usar essas informações para ajudar a identificar anomalias no tráfego de rede e para identificar um vetor de ataque específico. Por exemplo, a maioria dos ataques de reflexão UDP tem portas de origem específicas, como a porta de origem 53 para reflexão de DNS. Essa é uma assinatura clara de ataque que você pode identificar no registro do log de fluxo. Em resposta, você pode optar por

bloquear a porta de origem específica no nível da instância ou criar uma regra de ACL de rede para bloquear todo o protocolo se a aplicação não precisar dele.

Para obter mais informações sobre como usar os logs de fluxo da VPC para identificar anomalias de rede e vetores de ataque de DDoS, consulte Logs de fluxo da VPC e Logs de fluxo da VPC: registrar e visualizar fluxos de tráfego de rede.

Gerenciamento de visibilidade e proteção em várias contas

Em cenários em que você opera em várias contas da AWS e tem vários componentes para proteger, o uso de técnicas que permitem operar em escala e reduzir a sobrecarga operacional aumenta seus recursos de mitigação. Ao gerenciar recursos protegidos do AWS Shield Advanced em várias contas, você pode configurar o monitoramento centralizado usando o AWS Firewall Manager e o AWS Security Hub. Com o Firewall Manager, você pode criar uma política de segurança que impõe a conformidade da proteção contra DDoS em todas as suas contas. Você pode usar esses dois serviços juntos para gerenciar seus recursos protegidos em várias contas e centralizar o monitoramento desses recursos.

O Security Hub integra-se automaticamente ao Firewall Manager, permitindo que os clientes do Shield Advanced visualizem as descobertas de segurança em um único painel, juntamente com outros alertas de segurança de alta prioridade e status de conformidade. Por exemplo, quando o Shield Advanced detecta tráfego anômalo destinado a um recurso protegido em qualquer conta da AWS dentro do escopo, essa descoberta ficará visível no console do Security Hub. Se configurado, o Firewall Manager pode colocar automaticamente o recurso em conformidade, criando-o como um recurso protegido pelo Shield Advanced e, em seguida, atualizar o Security Hub quando o recurso estiver em um estado compatível.

Para obter mais informações sobre o monitoramento central de recursos protegidos do Shield, consulte Configurar monitoramento centralizado para eventos DDoS e remediar automaticamente recursos fora de conformidade.

Suporte

Se você enfrentar um ataque, também poderá se beneficiar do suporte da AWS ao avaliar a ameaça e revisar a arquitetura da sua aplicação ou solicitar outra assistência. É importante criar um plano de resposta para ataques de DDoS antes de um evento real. As práticas recomendadas descritas neste documento são medidas proativas que você implementa antes de iniciar uma aplicação, mas

ataques de DDoS contra sua aplicação ainda podem ocorrer. Analise as opções nesta seção para determinar os recursos de suporte mais adequados ao seu cenário. Sua equipe de conta pode avaliar seu caso de uso e aplicação e ajudar com perguntas ou desafios específicos que você tenha.

Se você estiver executando workloads de produção na AWS, considere assinar o Business Support, que fornece acesso 24 horas por dia, 7 dias por semana aos engenheiros de suporte na nuvem que podem ajudar com problemas de ataque de DDoS. Se você estiver executando workloads essenciais à missão, considere o Enterprise Support, que oferece a capacidade de abrir casos críticos e receber a resposta mais rápida de um engenheiro sênior de suporte na nuvem.

Se você estiver inscrito no AWS Shield Advanced e também estiver inscrito no Business Support ou no Enterprise Support, poderá configurar o engajamento proativo do Shield. Ele permite que você configure verificações de integridade, associe seus recursos e forneça informações de contato de operações 24 horas por dia, 7 dias por semana. Quando o Shield detecta sinais de DDoS e as verificações de integridade da aplicação mostram sinais de degradação, o AWS SRT entrará em contato com você de forma proativa. Esse é o nosso modelo de engajamento recomendado porque permite os tempos de resposta mais rápidos do AWS SRT e permite que o AWS SRT comece a solucionar problemas antes mesmo que o contato seja estabelecido com você.

O recurso de engajamento proativo exige que você configure uma verificação de integridade do Route 53 que meça com precisão a integridade da sua aplicação e esteja associada ao recurso protegido pelo Shield Advanced. Depois que uma verificação de integridade do Route 53 for associada ao console do Shield, o sistema de detecção Shield Advanced usará o status da verificação de integridade como um indicador da integridade da sua aplicação. O recurso de detecção baseada em integridade do Shield Advanced garantirá que você seja notificado e que as atenuações sejam colocadas mais rapidamente quando a aplicação não estiver íntegra. O AWS SRT entrará em contato com você para solucionar se a aplicação não íntegra está sendo alvo de um ataque de DDoS e colocar mitigações adicionais conforme necessário.

Concluir a configuração do compromisso proativo inclui adicionar detalhes de contato no console do Shield. O AWS SRT usará essas informações para entrar em contato com você. Você pode configurar até 10 contatos e fornecer notas adicionais se tiver requisitos ou preferências de contato específicos. Os contatos de engajamento proativo devem ter uma função 24 horas por dia, 7 dias por semana, como um centro de operações de segurança ou um indivíduo que esteja imediatamente disponível.

Você pode habilitar o engajamento proativo para todos os recursos ou para selecionar os principais recursos de produção em que o tempo de resposta é fundamental. Isso é feito atribuindo verificações de integridade somente a esses recursos.

Você também pode escalar para o AWS SRT criando um caso do AWS Support usando o console do AWS Support ou a API do Support se tiver um evento relacionado a DDoS que afete a disponibilidade da sua aplicação.

Conclusão

As práticas recomendadas descritas neste documento podem ajudar você a criar uma arquitetura resistente a DDoS que proteja a disponibilidade do sua aplicação, impedindo muitos ataques de DDoS comuns de infraestrutura e camada de aplicação. Até que ponto você segue essas práticas recomendadas ao arquitetar sua aplicação influenciará o tipo, o vetor e o volume de ataques de DDoS que você poderá mitigar. Você pode incorporar resiliência sem assinar um serviço de mitigação de DDoS. Ao optar por se inscrever no AWS Shield Advanced, você obterá recursos adicionais de suporte, visibilidade, mitigação e proteção de custos que protegem ainda mais uma arquitetura de aplicação já resiliente.

Colaboradores

Os colaboradores desse documento incluem:

- Jeffrey Lyon, proteção de perímetro da AWS
- Rodrigo Ferroni, especialista em segurança da AWS TAM
- Dmitriy Novikov, arquiteto de soluções da AWS
- Achraf Souk, arquiteto de soluções da AWS
- Yoshihisa Nakatani, arquiteto de soluções da AWS

Recursos

Outras fontes de leitura:

- 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 e 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

Revisões do documento

Para ser notificado sobre atualizações deste whitepaper, inscreva-se no RSS feed.

update-history-change	update-history-description	update-history-date
Atualização do whitepaper	Atualizado para incluir recomendações e recursos mais recentes. O AWS Global Accelerator foi adicionado como parte da proteção abrangente na borda. AWS Firewall Manager para monitoramento centralizado de eventos DDoS e correção automática de recursos fora de conformidade.	21 de setembro de 2021
Atualização do whitepaper	Atualizado para esclarecer a quebra de cache na seção Detectar e filtrar solicitações da web maliciosas (BP1, BP2) e o uso de ELB e ALB na seção Escalar para absorver (BP6). Diagramas atualizados e a Tabela 2, marcada como “Opção da região”. como BP8. Seção BP7 atualizada com mais detalhes.	18 de dezembro de 2019
Atualização do whitepaper	Atualizado para incluir o registro em log do AWS WAF como uma prática recomendada.	1º de dezembro de 2018
Atualização do whitepaper	Atualizado para incluir o AWS Shield, os recursos do	1º de junho de 2018

	AWS WAF, o AWS Firewall Manager e as práticas recomendadas relacionadas.	
Atualização do whitepaper	Adição de orientação de arquitetura prescritiva e atualizada para incluir o AWS WAF.	1º de junho de 2016
Publicação inicial	Whitepaper publicado.	1º de junho de 2015

Avisos

Os clientes são responsáveis por fazer sua própria avaliação independente das informações neste documento. Este documento é: (a) fornecido apenas para fins informativos, (b) representa as ofertas e práticas de produtos atuais da AWS, que estão sujeitas a alterações sem aviso prévio e (c) não cria nenhum compromisso ou garantia da AWS e suas afiliadas, fornecedores ou licenciadores. Os produtos ou serviços da AWS são fornecidos no “estado em que se encontram”, sem garantias, declarações ou condições de qualquer tipo, explícitas ou implícitas. As responsabilidades e obrigações da AWS com seus clientes são regidas por contratos da AWS, e este documento não modifica nem faz parte de nenhum contrato entre a AWS e seus clientes.