Livre blanc AWS

Bonnes pratiques AWS pour la résistance aux attaques DDoS

Bonnes pratiques AWS pour la résistance aux attaques DDoS: Livre blanc AWS

Table of Contents

Bonnes pratiques AWS pour la résilience DDoS

Date de publication : 21 septembre 2021 (Révisions du document)

Résumé

Il est important de protéger votre entreprise contre l'impact des attaques par déni de service distribué (DDoS), ainsi que des autres cyberattaques. Préserver la confiance des clients dans votre service en maintenant la disponibilité et la réactivité de votre application est une priorité absolue. Vous voulez également éviter des coûts directs inutiles lorsque votre infrastructure doit évoluer en réponse à une attaque. Amazon Web Services (AWS) s'engage à vous fournir les outils, les bonnes pratiques et les services nécessaires pour vous défendre contre les acteurs malveillants sur Internet. L'utilisation des bons services d'AWS permet de garantir une disponibilité, une sécurité et une résilience élevées.

Dans ce livre blanc, AWS fournit des conseils normatifs en matière de DDoS pour améliorer la résilience des applications exécutées sur AWS. Cela inclut une architecture de référence résiliente aux attaques DDoS qui peut être utilisée comme guide pour aider à protéger la disponibilité des applications. Ce livre blanc décrit également différents types d'attaques, tels que les attaques de la couche d'infrastructure et les attaques de la couche application. AWS explique quelles sont les bonnes pratiques pour gérer chaque type d'attaque. En outre, les services et les fonctions qui s'inscrivent dans une stratégie d'atténuation des attaques DDoS sont décrits et la façon dont chacun peut être utilisé pour protéger vos applications est expliquée.

Ce document est destiné aux décideurs informatiques et aux ingénieurs en sécurité qui connaissent les concepts de base de la mise en réseau, de la sécurité et d'AWS. Chaque section contient des liens vers de la documentation AWS qui fournit plus de détails sur la bonne pratique ou capacité.

Introduction : attaques par déni de service

Une attaque par déni de service (DoS) est une tentative délibérée de rendre un site web ou une application indisponible pour les utilisateurs, par exemple en l'inondant de trafic réseau. Les attaquants utilisent diverses techniques qui consomment de grandes quantités de bande passante réseau ou bloquent d'autres ressources système, perturbant ainsi l'accès des utilisateurs légitimes. Dans sa forme la plus simple, un attaquant isolé utilise une source unique pour mener une attaque DoS contre une cible, comme le montre l'image suivante.

Tableau 1 : Diagramme de l'attaque DoS

Lors d'une attaque DDoS, un attaquant utilise plusieurs sources pour orchestrer une attaque contre une cible. Ces sources peuvent inclure des groupes distribués d'ordinateurs, de routeurs, d'appareils IoT et d'autres points de terminaison infectés par des logiciels malveillants. Le diagramme suivant montre qu'un réseau d'hôtes compromis participe à l'attaque, générant un flot de paquets ou de demandes visant à submerger la cible.

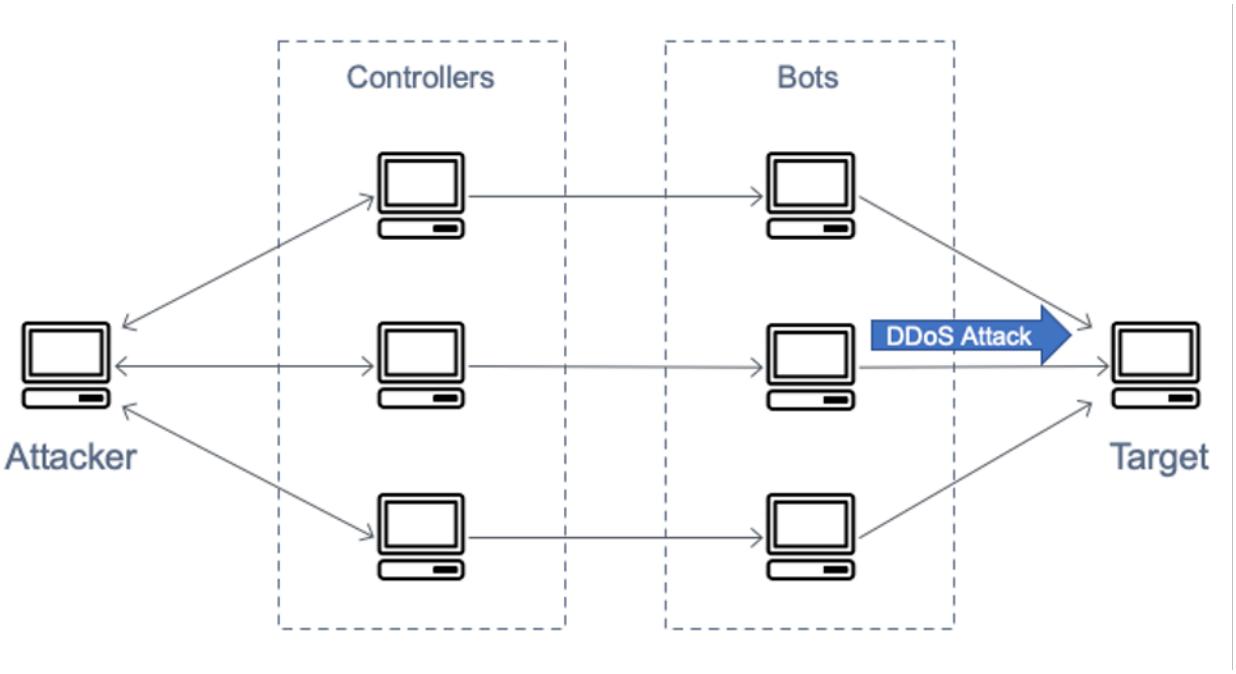

Diagramme de l'attaque DDoS

Le modèle d'interconnexion de systèmes ouverts (OSI) comprend sept couches et elles sont décrites dans le tableau Modèle d'interconnexion de systèmes ouverts (OSI). Les attaques DDoS sont les plus courantes aux niveaux trois, quatre, six et sept. Les attaques des couches trois et quatre

correspondent aux couches réseau et transport du modèle OSI. Dans ce document, AWS les désigne collectivement comme des attaques de la couche d'infrastructure. Les attaques des couches six et sept correspondent aux couches Presentation et Application du modèle OSI. AWS traitera ces problèmes ensemble en tant qu'attaques de la couche application. Des exemples de ces types d'attaques sont présentés dans les sections suivantes.

Modèle d'interconnexion de systèmes ouverts (OSI)

#	Couche	Unité	Description	Exemples Vector
7	Application	Données	Processus réseau en direction de l'application	Flux HTTP, Flux de requête DNS
6	Présentation	Données	Représentation et chiffrement des données	Abus de TLS
5	Session	Données	Communication inter-hôte	S/O
4	Transport	Segments	Connexions de bout en bout et fiabilité	Flux SYN
3	Réseau	Paquets	Détermination de chemin et adressage logique	Attaques par réflexion basées sur des protocoles UDP
2	Liaison de données	Images	Adressage physique	S/O
1	Physique	Bits	Transmission multimédia, de signal et binaire	S/O

Rubriques

Attaques de couche d'infrastructure

Les attaques DDoS les plus courantes, les attaques par réflexion UDP (User Datagram Protocol) et les inondations de synchronisation (SYN), sont des attaques de la couche d'infrastructure. Un attaquant peut utiliser l'une ou l'autre de ces méthodes pour générer d'importants volumes de trafic susceptibles d'inonder la capacité d'un réseau ou d'immobiliser des ressources sur des systèmes tels que des serveurs, des pare-feu, un système de prévention des intrusions (IPS) ou un équilibreur de charge. Bien que ces attaques puissent être faciles à identifier, pour les atténuer efficacement, vous devez disposer d'un réseau ou de systèmes qui augmentent la capacité plus rapidement que le flux de trafic entrant. Cette capacité supplémentaire est nécessaire pour filtrer ou absorber le trafic d'attaque, libérant ainsi le système et l'application pour répondre au trafic client légitime.

Rubriques

Attaques par réflexion basées sur des protocoles UDP

Les attaques par réflexion UDP (User Datagram Protocol) exploitent le fait qu'UDP soit un protocole sans état. Les attaquants peuvent créer un paquet de demande UDP valide répertoriant l'adresse IP de la cible d'attaque en tant qu'adresse IP source UDP. L'attaquant a maintenant falsifié (usurpé) l'adresse IP source du paquet de requête UDP. Le paquet UDP contient l'adresse IP source usurpée et est envoyé par l'attaquant à un serveur intermédiaire. Le serveur est amené à envoyer ses paquets de réponse UDP à l'adresse IP ciblée de la victime plutôt qu'à l'adresse IP de l'attaquant. Le serveur intermédiaire est utilisé car il génère une réponse plusieurs fois supérieure au paquet de demande, ce qui amplifie efficacement la quantité de trafic d'attaque envoyée à l'adresse IP cible.

Le facteur d'amplification est le rapport entre la taille de la réponse et la taille de la demande, et il varie en fonction du protocole utilisé par l'attaquant : DNS, NTP, SSDP, CLDAP, Memcached, ChargeN ou QOTD. Par exemple, le facteur d'amplification pour le DNS peut être de 28 à 54 fois le nombre d'octets d'origine. Par conséquent, si un attaquant envoie une charge utile de demande de 64 octets à un serveur DNS, il peut générer plus de 3 400 octets de trafic indésirable vers une cible

d'attaque. Les attaques par réflexion UDP sont responsables d'un volume de trafic plus important que les autres attaques. La figure UDP Reflection Attack illustre la tactique de réflexion et l'effet d'amplification.

Attaque par réflexion basée sur des protocoles UDP

Attaques par inondation SYN

Lorsqu'un utilisateur se connecte à un service TCP (Transmission Control Protocol), tel qu'un serveur web, son client envoie un paquet de synchronisation SYN. Le serveur renvoie un paquet SYN-ACK en accusé de réception, et finalement le client répond avec un paquet d'accusé de réception (ACK), qui termine la liaison à trois voies attendue. L'image suivante illustre cette liaison type.

Liaison à 3 voies SYN

Lors d'une attaque par inondation SYN, un client malveillant envoie un grand nombre de paquets SYN, mais n'envoie jamais les derniers paquets ACK pour terminer les liaisons. Le serveur est laissé en attente d'une réponse aux connexions TCP semi-ouvertes et finit par manquer de capacité pour accepter de nouvelles connexions TCP. Cela peut empêcher de nouveaux utilisateurs de se connecter au serveur. L'attaque tente d'immobiliser les connexions au serveur disponibles afin que les ressources ne soient pas disponibles pour les connexions légitimes. Bien que les inondations SYN puissent atteindre des centaines de Gbit/s, l'objectif de l'attaque n'est pas d'augmenter le volume du trafic SYN.

Attaques de couche d'application

Un attaquant peut cibler l'application elle-même en utilisant une attaque de couche 7 ou de couche application. Dans ces attaques, similaires aux attaques d'infrastructure SYN flood, l'attaquant tente de surcharger des fonctions spécifiques d'une application pour la rendre indisponible ou ne pas répondre aux utilisateurs légitimes. Cela peut parfois être réalisé avec de très faibles volumes de demandes qui ne génèrent qu'un faible volume de trafic réseau. Cela peut rendre l'attaque difficile à détecter et à atténuer. Parmi les exemples d'attaques de la couche application, citons les inondations HTTP, les attaques anti-cache et les inondations XML-RPC de WordPress.

Lors d'une attaque par inondation HTTP, un attaquant envoie des requêtes HTTP qui semblent provenir d'un utilisateur valide de l'application web. Certaines inondations HTTP ciblent une

ressource spécifique, tandis que des inondations HTTP plus complexes tentent d'émuler l'interaction humaine avec l'application. Cela peut augmenter la difficulté d'utiliser des techniques d'atténuation courantes telles que la limitation du taux de demande.

Les attaques par destruction de cache sont un type d'inondation HTTP qui utilise des variations dans la chaîne de requête pour contourner la mise en cache du réseau de diffusion de contenu (CDN). Au lieu de pouvoir renvoyer des résultats mis en cache, le CDN doit contacter le serveur d'origine pour chaque demande de page, et ces récupérations d'origine entraînent une charge supplémentaire sur le serveur web de l'application.

Avec une attaque d'inondation WordPress XML-RPC, également connue sous le nom d'inondation de pingback WordPress, un attaquant cible un site web hébergé sur le logiciel de gestion de contenu WordPress. L'attaquant utilise à mauvais escient la fonction de l'API XML-RPC pour générer un flot de requêtes HTTP. La fonction pingback permet à un site web hébergé sur WordPress (site A) d'avertir un autre site WordPress (site B) via un lien que le site A a créé vers le site B. Le site B tente ensuite de récupérer le site A pour vérifier l'existence du lien. Dans un flux de pingback, l'attaquant utilise cette capacité à mauvais escient pour amener le site B à attaquer le site A. Ce type d'attaque a une signature claire : WordPress est généralement présent dans l'agent utilisateur de l'en-tête de la requête HTTP.

Il existe d'autres formes de trafic malveillant qui peuvent avoir un impact sur la disponibilité d'une application. Les robots Scraper automatisent les tentatives d'accès à une application web pour voler du contenu ou enregistrer des informations concurrentielles, telles que les prix. Les attaques par force brute et par « credential stuffing » sont des efforts programmés visant à obtenir un accès non autorisé aux zones sécurisées d'une application. Il ne s'agit pas uniquement d'attaques DDoS ; mais leur nature automatisée peut ressembler à une attaque DDoS et elles peuvent être atténuées en mettant en œuvre certaines des bonnes pratiques abordées dans ce document.

Les attaques de la couche application peuvent également cibler les services DNS (Domain Name System). La plus courante de ces attaques est une inondation de requêtes DNS au cours de laquelle un attaquant utilise de nombreuses requêtes DNS bien formées pour épuiser les ressources d'un serveur DNS. Ces attaques peuvent également inclure un composant briseur de cache dans lequel l'attaquant randomise la chaîne de sous-domaine pour contourner le cache DNS local d'un résolveur donné. Par conséquent, le résolveur ne peut pas tirer parti des requêtes de domaine mises en cache et doit contacter à plusieurs reprises le serveur DNS faisant autorité, ce qui amplifie l'attaque.

Si une application web est mise à disposition via le protocole TLS (Transport Layer Security), un attaquant peut également choisir d'attaquer le processus de négociation TLS. Le protocole TLS coûte cher en termes de calcul, de sorte qu'un attaquant, en générant une charge de travail supplémentaire

sur le serveur pour traiter des données illisibles (ou inintelligibles (texte chiffré)) en tant que liaison légitime, peut réduire la disponibilité du serveur. Dans une variante de cette attaque, un attaquant termine l'établissement de liaison TLS, mais renégocie perpétuellement la méthode de chiffrement. Un attaquant peut également tenter d'épuiser les ressources du serveur en ouvrant et en fermant de nombreuses sessions TLS.

Techniques d'atténuation

Certaines formes d'atténuation des attaques DDoS sont automatiquement incluses dans les services AWS. La résilience DDoS peut être encore améliorée en utilisant une architecture AWS avec des services spécifiques, abordés dans les sections suivantes, et en mettant en œuvre des bonnes pratiques supplémentaires pour chaque partie du flux réseau entre les utilisateurs et votre application.

Tous les clients AWS bénéficient des protections automatiques d'AWS Shield Standard, sans frais supplémentaires. AWS Shield Standard protège contre les attaques DDoS les plus fréquentes de la couche réseau et de transport qui ciblent les sites web et les applications. Cette protection est toujours activée, préconfigurée, statique, et ne fournit aucun rapport ni aucune analytique. Elle est proposée sur tous les services AWS et dans toutes les régions AWS. Dans les régions AWS, les attaques DDoS sont détectées et le système Shield Standard définit automatiquement la référence du trafic, identifie les anomalies et, si nécessaire, crée des mesures d'atténuation. Vous pouvez utiliser AWS Shield Standard dans le cadre d'une architecture résiliente aux attaques DDoS pour protéger les applications web et non web.

Vous pouvez également utiliser des services AWS qui opèrent à partir d'emplacements périphériques, tels qu'Amazon CloudFront, Global Accelerator et Route 53 pour créer une protection de disponibilité complète contre toutes les attaques connues de la couche d'infrastructure. Ces services font partie du réseau AWS Global Edge et peuvent améliorer la résilience DDoS de votre application lors de la distribution de tout type de trafic d'application à partir d'emplacements périphériques répartis dans le monde entier. Vous pouvez exécuter votre application dans n'importe quelle région AWS et utiliser ces services pour protéger la disponibilité de votre application et optimiser les performances de votre application pour les utilisateurs finaux légitimes.

Les avantages liés à l'utilisation d'Amazon CloudFront, de Global Accelerator et d'Amazon Route 53 sont les suivants :

- Accès à Internet et à la capacité d'atténuation des attaques DDoS sur l'ensemble du réseau AWS Global Edge. Ceci est utile pour atténuer les attaques volumétriques plus importantes, qui peuvent atteindre l'échelle du térabit.
- Les systèmes d'atténuation des attaques DDoS AWS Shield sont intégrés aux services AWS en périphérie, ce qui réduit le temps d'atténuation de quelques minutes à moins d'une seconde.
- Les techniques d'atténuation des inondations SYN sans état utilisent un proxy et vérifient les connexions entrantes avant de les transmettre au service protégé. Cela garantit que seules les

connexions valides atteignent votre application, tout en protégeant vos utilisateurs finaux légitimes contre les baisses de faux positifs.

- Systèmes automatiques d'ingénierie du trafic qui dispersent ou isolent l'impact des attaques DDoS volumétriques de grande envergure. Tous ces services isolent les attaques à la source avant qu'elles n'atteignent votre point d'origine, ce qui réduit l'impact sur les systèmes protégés par ces services.
- La défense de la couche application, lorsqu'elle est combinée à AWS WAF qui ne nécessite pas de modification de l'architecture d'application actuelle (par exemple, dans une région AWS ou un centre de données sur site).

Il n'y a aucun frais pour le transfert de données entrant activé sur AWS et vous ne payez pas pour le trafic d'attaque DDoS atténué par AWS Shield. Le diagramme d'architecture suivant inclut les services AWS Global Edge Network.

Cette architecture inclut plusieurs services AWS qui peuvent vous aider à améliorer la résilience de votre application web contre les attaques DDoS. Le tableau Résumé des bonnes pratiques fournit un résumé de ces services et des fonctions qu'ils peuvent fournir. AWS a labelisé chaque service avec un indicateur de bonnes pratiques (BP1, BP2) pour faciliter la consultation dans ce document. Par exemple, une section à venir traite des fonctionnalités fournies par Amazon CloudFront et Global Accelerator qui incluent l'indicateur de bonnes pratiques BP1.

Tableau 2 - Résumé des bonnes pratiques

AWS Edge	Région AWS					
	Utilisation d'Amazon Cl t (BP1) avec AWS WAF (BP2)	Utilisation de Global Accelerator (BP1)	Utilisation d'Amazon Rc	Utilisation d'Elastic Load Balancing (BP6) avec AWS WAF (BP2)	Utilisation de groupes de sécurité et de listes de contrôle d'accès réseau dans	Utilisation d'Amazon EC2 Auto Scaling (BP7)

AWS Edge	Région AWS					
					Amazon VPC (BP5)	
Atténuation des attaques de couche 3 (par exemple, réflexion UDP)	✓	✓	✓	✓	✓	✓
Atténuation des attaques de couche 4 (par exemple, inondation SYN)	✓	✓	✓	✓		
Atténuation des attaques de couche 6 (par exemple, TLS)	✓	✓	✓	✓		
Réduire la surface d'attaque	✓	✓	✓	✓	✓	

AWS Edge	Région AWS					
Mise à l'échelle pour absorber le trafic de la couche	✓	✓	✓	✓	✓	✓
Atténuation des attaques de couche 7 (couche d'application)	✓	✓	✓	✓	✓	✓
Isolation géographique et dispersion du trafic excédentaire et des attaques DDoS plus importantes	✓	✓	✓			

AWS Edge	Région AWS					
✓ (*) : en cas d'utilisation avec AWS WAF et l'Application Load Balancer						

Vous pouvez également vous préparer à réagir aux attaques DDoS et à les atténuer en vous abonnant à AWS Shield Advanced.

Les clients bénéficient d'une détection personnalisée basée sur :

- Les modèles de trafic spécifiques de votre application.
- la protection contre les attaques DDoS de couche 7, y compris AWS WAF sans frais supplémentaires.
- L'accès à un support spécialisé 24 h/24, 7 j/7 à partir d'AWS SRT.
- La gestion centralisée des politiques de sécurité via AWS Firewall Manager.
- La protection des coûts pour se prémunir contre les frais de mise à l'échelle résultant de pics d'utilisation liés aux attaques DDoS.

Ce service facultatif d'atténuation des attaques DDoS permet de protéger les applications hébergées dans toutes les régions AWS. Le service est disponible dans le monde entier pour CloudFront, Route 53 et Global Accelerator. L'utilisation de Shield Advanced avec des adresses IP Elastic vous permet de protéger les instances Network Load Balancer (NLB) ou Amazon EC2.

Les avantages de l'utilisation d'AWS Shield Advanced incluent :

- Accès à AWS SRT afin d'obtenir de l'aide pour atténuer les attaques DDoS qui ont un impact sur la disponibilité des applications.
- Visibilité des attaques DDoS à l'aide des métriques et des alarmes d'AWS Management Console, de l'API et d'Amazon CloudWatch.
- Accès à l'historique de tous les événements DDoS des 13 derniers mois.

- Accès au pare-feu d'application web AWS (AWS WAF), sans frais supplémentaires pour atténuer les attaques DDoS de la couche application (en cas d'utilisation avec Amazon CloudFront ou Application Load Balancer).
- Établissement de référence automatique des attributs de trafic web, en cas d'utilisation avec AWS WAF.
- Accès à AWS Firewall Manager, sans frais supplémentaires, pour l'application automatisée des politiques.
- Seuils de détection sensibles qui acheminent le trafic vers le système d'atténuation des attaques DDoS plus tôt et peuvent améliorer le délai d'atténuation des attaques contre Amazon EC2 ou Network Load Balancer, lorsqu'ils sont utilisés avec une adresse IP Elastic.
- Protection des coûts qui vous permet de demander un remboursement limité des coûts liés à la mise à l'échelle résultant d'une attaque DDoS.
- Contrat de niveau de service amélioré spécifique aux clients AWS Shield Advanced.
- Engagement proactif de la part d'AWS SRT lorsqu'un événement Shield est détecté.
- Des groupes de protection qui vous permettent de regrouper des ressources, offrant ainsi un moyen en libre-service de personnaliser la portée de détection et d'atténuation de votre application en traitant plusieurs ressources comme une seule unité. Le regroupement des ressources améliore la précision de la détection, réduit les faux positifs, facilite la protection automatique des ressources nouvellement créées et accélère le temps nécessaire pour atténuer les attaques contre de nombreuses ressources qui constituent une seule application. Pour plus d'informations sur les groupes de protection, consultez Groupes de protection Shield Advanced.

Pour obtenir la liste complète des fonctions AWS Shield Advanced et pour plus d'informations sur AWS Shield, consultez Fonctionnement d'AWS Shield.

Rubriques

Bonnes pratiques pour l'atténuation des attaques DDoS

Dans les sections suivantes, chacune des meilleures pratiques recommandées pour l'atténuation des attaques DDoS est décrite plus en détail. Pour obtenir un guide rapide et facile à mettre en œuvre

sur la création d'une couche d'atténuation des attaques DDoS pour les applications web statiques ou dynamiques, consultez Comment protéger les applications web dynamiques contre les attaques DDoS.

Défense de la couche d'infrastructure (BP1, BP3, BP6, BP7)

Dans un environnement de centre de données classique, vous pouvez atténuer les attaques DDoS de la couche d'infrastructure en utilisant des techniques telles que le surprovisionnement de capacité, le déploiement de systèmes d'atténuation des attaques DDoS ou le nettoyage du trafic à l'aide de services d'atténuation des attaques DDoS. SurAWS, les fonctionnalités d'atténuation des attaques DDoS sont automatiquement fournies, mais vous pouvez optimiser la résilience DDoS de votre application en faisant des choix d'architecture qui tirent le meilleur parti de ces fonctionnalités et vous permettent également d'évoluer en fonction du trafic excédentaire.

Les principaux facteurs à prendre en compte pour atténuer les attaques DDoS volumétriques incluent la garantie d'une capacité et d'une diversité de transit suffisantes et la protection des ressources AWS, telles que les instances Amazon EC2, contre le trafic d'attaque.

Certains types d'instances Amazon EC2 prennent en charge des fonctionnalités qui peuvent gérer plus facilement d'importants volumes de trafic, par exemple des interfaces de bande passante réseau allant jusqu'à 100 Gbit/s et une mise en réseau améliorée. Cela permet d'éviter la congestion de l'interface pour le trafic qui a atteint l'instance Amazon EC2. Les instances qui prennent en charge la mise en réseau améliorée offrent des performances d'E/S, une bande passante supérieure et une utilisation du processeur inférieure par rapport aux implémentations traditionnelles. Cela améliore la capacité de l'instance à gérer de grands volumes de trafic et, en fin de compte, la rend hautement résiliente face à la charge de paquets par seconde (pps).

Pour permettre ce niveau élevé de résilience, AWS recommande d'utiliser des instances dédiées Amazon EC2 ou des instances Amazon EC2 avec un débit réseau plus élevé qui ont un suffixe N et la prise en charge de la mise en réseau améliorée avec jusqu'à 100 Gbit/s de bande passante réseau, par exemple des instances c6gn.16xlarge et c5n.18xlarge ou matériel nu (telles que c5n.metal).

Pour plus d'informations sur les instances Amazon EC2 qui prennent en charge les interfaces réseau 100 Gigabit et la mise en réseau améliorée, consultez Types d'instances Amazon EC2.

Le module requis pour une mise en réseau améliorée et le jeu d'attributs enaSupport requis sont inclus avec Amazon Linux 2 et les dernières versions de l'AMI Amazon Linux. Par conséquent, si vous lancez une instance avec la dernière version HVM d'Amazon Linux sur un type d'instance

pris en charge, la mise en réseau améliorée est déjà activée pour votre instance. Pour plus d'informations, consultez Tester si la mise en réseau améliorée est activée. Pour plus d'informations sur la façon d'activer la mise en réseau améliorée, consultez Mise en réseau améliorée sous Linux.

Amazon EC2 avec Auto Scaling (BP7)

Un autre moyen d'atténuer à la fois les attaques de l'infrastructure et de la couche application consiste à opérer à grande échelle. Si vous avez des applications web, vous pouvez utiliser des équilibreurs de charge pour distribuer le trafic vers un certain nombre d'instances Amazon EC2 surprovisionnées ou configurées pour une mise à l'échelle automatique. Ces instances peuvent gérer des pics de trafic soudains qui surviennent pour n'importe quelle raison, y compris une foule éclair ou une attaque DDoS de la couche application. Vous pouvez définir des alarmes Amazon CloudWatch pour lancer Auto Scaling afin de dimensionner automatiquement la taille de votre flotte Amazon EC2 en réponse aux événements que vous définissez, tels que le processeur, la RAM, les E/S réseau et même les métriques personnalisées. Cette approche protège la disponibilité des applications en cas d'augmentation inattendue du volume de demandes. Lorsque vous utilisez Amazon CloudFront, un Application Load Balancer, des Classic Load Balancers ou un Network Load Balancer avec votre application, la négociation TLS est gérée par la distribution (Amazon CloudFront) ou par l'équilibreur de charge. Ces fonctions aident à protéger vos instances contre les attaques basées sur TLS en les adaptant pour gérer les demandes légitimes et les attaques d'abus TLS.

Pour plus d'informations sur l'utilisation d'Amazon CloudWatch pour appeler Auto Scaling, consultez Surveillance des métriques Amazon CloudWatch pour vos groupes et instances Auto Scaling.

Amazon EC2 fournit une capacité de calcul redimensionnable afin que vous puissiez rapidement augmenter ou réduire vos besoins en fonction de l'évolution des besoins. Vous pouvez effectuer une mise à l'échelle horizontale en ajoutant automatiquement des instances à votre application en mettant à l'échelle la taille de votre groupe Amazon EC2 Auto Scaling et vous pouvez effectuer une mise à l'échelle verticale en utilisant des types d'instances EC2 plus importants.

Elastic Load Balancing (BP6)

Les attaques DDoS importantes peuvent dépasser la capacité d'une seule instance Amazon EC2. Avec Elastic Load Balancing (ELB), vous pouvez réduire le risque de surcharge de votre application en répartissant le trafic entre de nombreuses instances de backend. Elastic Load Balancing peut évoluer automatiquement, ce qui vous permet de gérer des volumes plus importants lorsque vous avez un trafic supplémentaire imprévu, par exemple en raison de foules flash ou d'attaques DDoS. Pour les applications créées au sein d'un Amazon VPC, il existe trois types d'ELB à prendre en

compte, en fonction de votre type d'application : Application Load Balancer (ALB), Classic Load Balancer (CLB) et Network Load Balancer (NLB).

Pour les applications web, vous pouvez utiliser l'Application Load Balancer pour acheminer le trafic en fonction du contenu et n'accepter que des demandes web bien formées. L'Application Load Balancer bloque de nombreuses attaques DDoS courantes, telles que les inondations de synchronisation ou les attaques par réflexion UDP, protégeant ainsi votre application contre l'attaque. L'Application Load Balancer se met automatiquement à l'échelle pour absorber le trafic supplémentaire lorsque ces types d'attaques sont détectés. Les activités de mise à l'échelle dues aux attaques de la couche d'infrastructure sont transparentes pour les clients AWS et n'affectent pas votre facture.

Pour plus d'informations sur la protection des applications web avec l'Application Load Balancer, consultez Mise en route avec les Application Load Balancers.

Pour les applications basées sur TCP, vous pouvez utiliser l'équilibreur de charge réseau pour acheminer le trafic vers des cibles (par exemple, des instances Amazon EC2) avec une latence ultra-faible. L'un des principaux aspects du Network Load Balancer est que tout trafic qui atteint l'équilibreur de charge sur un écouteur valide sera acheminé vers vos cibles, et non absorbé. Vous pouvez utiliser Shield Advanced pour configurer la protection DDoS pour les adresses IP Elastic. Lorsqu'une adresse IP Elastic est attribuée par zone de disponibilité au Network Load Balancer, Shield Advanced applique les protections DDoS appropriées pour le trafic du Network Load Balancer.

Pour plus d'informations sur la protection des applications TCP avec le Network Load Balancer, consultez Mise en route avec les Network Load Balancers.

Tirez parti des emplacements périphériques AWS à des fins d'évolutivité (BP1, BP3)

L'accès à des connexions Internet diversifiées et à grande échelle peut considérablement augmenter votre capacité à optimiser la latence et le débit pour les utilisateurs, à absorber les attaques DDoS et à isoler les pannes tout en minimisant l'impact sur la disponibilité de votre application. Les emplacements périphériques AWS fournissent une couche supplémentaire d'infrastructure réseau qui fournit ces avantages à toute application web qui utilise Amazon CloudFront, Global Accelerator et Amazon Route 53. Grâce à ces services, vous pouvez protéger complètement en périphérie vos applications exécutées à partir de régions AWS.

Livraison d'applications web à la périphérie (BP1)

Amazon CloudFront est un service qui peut être utilisé pour diffuser l'intégralité de votre site web, y compris du contenu statique, dynamique, en streaming et interactif. Les connexions persistantes et les paramètres de durée de vie variable (TTL) peuvent être utilisés pour décharger le trafic de votre origine, même si vous ne diffusez pas de contenu pouvant être mis en cache. L'utilisation de ces fonctions CloudFront réduit le nombre de demandes et de connexions TCP vers votre origine, contribuant ainsi à protéger votre application web contre les inondations HTTP. CloudFront n'accepte que les connexions bien formées, ce qui permet d'empêcher de nombreuses attaques DDoS courantes, telles que les inondations SYN et les attaques par réflexion UDP, d'atteindre votre origine. De plus, les attaques DDoS sont géographiquement isolées à proximité de la source, ce qui empêche le trafic d'affecter d'autres emplacements. Toutes ces fonctionnalités peuvent considérablement améliorer votre capacité à continuer à assurer le trafic en direction des utilisateurs pendant les attaques DDoS massives. Vous pouvez utiliser CloudFront pour protéger une origine sur AWS ou ailleurs sur Internet.

Si vous utilisez Amazon S3 pour diffuser du contenu statique sur Internet, AWS vous recommande d'utiliser Amazon CloudFront pour protéger votre compartiment. Vous pouvez utiliser l'identification d'accès à l'origine (OAI) pour garantir que les utilisateurs n'accèdent à vos objets qu'à l'aide des URL CloudFront.

Pour plus d'informations sur OAI, consultez Restriction de l'accès au contenu Amazon S3 à l'aide d'une identité d'accès à l'origine.

Pour plus d'informations sur la protection et l'optimisation des performances des applications web avec Amazon CloudFront, consultez Mise en route avec CloudFront.

Protégez le trafic réseau plus éloigné de votre origine à l'aide d'AWS Global Accelerator (BP1)

Global Accelerator est un service réseau qui améliore la disponibilité et les performances du trafic des utilisateurs jusqu'à 60 %. Pour ce faire, il suffit d'acheminer le trafic entrant au niveau de l'emplacement périphérique le plus proche de vos utilisateurs et de le router via l'infrastructure réseau mondiale d'AWS vers votre application, qu'elle s'exécute dans une ou plusieurs régions AWS.

Global Accelerator achemine le trafic TCP et UDP vers le point de terminaison optimal en fonction des performances de la région AWS la plus proche de l'utilisateur. En cas de défaillance d'une application, Global Accelerator fournit un basculement vers le meilleur point de terminaison suivant

dans les 30 secondes. Global Accelerator utilise la vaste capacité du réseau mondial AWS et les intégrations avec Shield, telles qu'une capacité de proxy SYN sans état qui remet en question les nouvelles tentatives de connexion et ne sert que les utilisateurs finaux légitimes, pour protéger les applications.

Vous pouvez mettre en œuvre une architecture résiliente aux attaques DDoS qui offre bon nombre des mêmes avantages que les bonnes pratiques de diffusion d'applications web en périphérie, même si votre application utilise des protocoles non pris en charge par CloudFront ou si vous exploitez une application web qui nécessite des adresses IP statiques globales. Par exemple, vous pouvez avoir besoin d'adresses IP que vos utilisateurs finaux peuvent ajouter à la liste d'autorisation de leurs pare-feu et qui ne sont utilisées par aucun autre client AWS. Dans ces scénarios, vous pouvez utiliser Global Accelerator pour protéger les applications web exécutées sur l'équilibreur de charge d'application et également AWS WAF pour détecter et atténuer les inondations de demandes de couche d'application web.

Pour plus d'informations sur la protection et l'optimisation des performances du trafic réseau à l'aide de Global Accelerator, consultez Premiers pas avec Global Accelerator.

Résolution des noms de domaine à la périphérie (BP3)

Amazon Route 53 est un service DNS (Domain Name System) hautement disponible et évolutif qui peut être utilisé pour diriger le trafic vers votre application web. Il comprend des fonctionnalités avancées telles que le flux de trafic, les vérifications et la surveillance de l'état, le routage basé sur la latence et le Geo DNS. Ces fonctions avancées vous permettent de contrôler la façon dont le service répond aux demandes DNS afin d'améliorer les performances de votre application web et d'éviter les pannes de site.

Amazon Route 53 utilise des techniques telles que le partitionnement aléatoire et l'entrelacement anycast, qui peuvent aider les utilisateurs à accéder à votre application même si le service DNS est la cible d'une attaque DDoS.

Avec le partitionnement aléatoire, chaque serveur de noms de votre ensemble de délégation correspond à un ensemble unique d'emplacements périphériques et de chemins Internet. Cela fournit une plus grande tolérance aux pannes et minimise les chevauchements entre les clients. Si un serveur de noms du jeu de délégation n'est pas disponible, les utilisateurs peuvent réessayer et recevoir une réponse d'un autre serveur de noms situé à un emplacement périphérique différent.

L'entrelacement anycast permet à chaque requête DNS d'être traitée par l'emplacement le plus optimal, répartissant ainsi la charge du réseau et réduisant la latence DNS. Cela fournit une réponse

plus rapide aux utilisateurs. En outre, Amazon Route 53 peut détecter des anomalies dans la source et le volume des requêtes DNS, et hiérarchiser les demandes des utilisateurs dont la fiabilité est reconnue.

Pour plus d'informations sur l'utilisation d'Amazon Route 53 pour acheminer les utilisateurs vers votre application, consultez Mise en route avec Amazon Route 53.

Défense de la couche application (BP1, BP2)

Bon nombre des techniques abordées jusqu'à présent dans ce document sont efficaces pour atténuer l'impact des attaques DDoS au niveau de l'infrastructure sur la disponibilité de votre application. Pour vous défendre également contre les attaques de la couche application, vous devez mettre en œuvre une architecture qui vous permet de détecter, de dimensionner pour absorber et bloquer spécifiquement les demandes malveillantes. Il s'agit d'une considération importante car les systèmes d'atténuation des attaques DDoS basés sur le réseau sont généralement inefficaces pour atténuer les attaques complexes de la couche applicative.

Détecter et filtrer les demandes web malveillantes avec (BP1, BP2)

Lorsque votre application est exécutée sur AWS, vous pouvez tirer parti d'Amazon CloudFront et d'AWS WAF pour vous aider à vous défendre contre les attaques DDoS de la couche application.

Amazon CloudFront vous permet de mettre en cache du contenu statique et de le diffuser à partir d'emplacements AWS périphériques, ce qui peut vous aider à réduire la charge sur votre origine. Cela peut également aider à réduire la charge du serveur en empêchant le trafic non web d'atteindre votre origine. En outre, CloudFront peut fermer automatiquement les connexions des attaquants en lecture lente ou en écriture lente (par exemple, Slowloris).

En utilisant AWS WAF, vous pouvez configurer des listes de contrôle d'accès web (ACL Web) sur vos distributions CloudFront ou vos Application Load Balancers pour filtrer et bloquer les demandes en fonction des signatures de demande. Chaque ACL web se compose de règles que vous pouvez configurer pour que la correspondance de chaîne ou d'expression régulière corresponde à un ou plusieurs attributs de demande, tels que l'URI (Uniform Resource Identifier), la chaîne de requête, la méthode HTTP ou la clé d'en-tête. En outre, en utilisant les règles basées sur les taux d'AWS WAF, vous pouvez bloquer automatiquement les adresses IP des mauvais acteurs lorsque les demandes correspondant à une règle dépassent un seuil que vous définissez.

Les demandes provenant d'adresses IP de clients incriminées recevront les réponses d'erreur 403 Interdit et resteront bloquées jusqu'à ce que le taux de demandes tombe en dessous du seuil.

Ceci est utile pour atténuer les attaques par inondation HTTP déguisées en trafic web normal. Pour bloquer les attaques basées sur la réputation de l'adresse IP, vous pouvez créer des règles à l'aide de conditions de correspondance d'adresses IP ou utiliser des règles gérées pour AWS WAF proposées par les vendeurs sur AWS Marketplace. AWS WAF propose directement les règles gérées par AWS en tant que service géré où vous pouvez choisir des groupes de règles de réputation IP. Le groupe de règles de la liste de réputation IP Amazon contient des règles basées sur les renseignements internes sur les menaces Amazon. Ceci est utile si vous souhaitez bloquer les adresses IP généralement associées à des robots ou à d'autres menaces. Le groupe de règles de la liste des IP anonymes contient des règles permettant de bloquer les demandes émanant des services qui permettent de masquer l'identité de l'utilisateur. Il s'agit notamment des demandes provenant de VPN, de proxys, de nœuds Tor et de plateformes cloud (y compris AWS). AWS WAF et CloudFront vous permettent également de définir des restrictions géographiques pour bloquer ou autoriser les demandes provenant de certains pays. Cela peut aider à bloquer les attaques provenant d'emplacements géographiques où vous ne comptez pas servir les utilisateurs.

Pour vous aider à identifier les demandes malveillantes, consultez les journaux de votre serveur web ou utilisez les fonctionnalités AWS WAF de journalisation et de demandes échantillonnées. En activant la journalisation AWS WAF, vous obtenez des informations détaillées sur le trafic analysé par l'ACL web. AWS WAF prend en charge le filtrage des journaux, ce qui vous permet de spécifier quelles demandes web sont enregistrées et quelles demandes sont ignorées du journal après l'inspection.

Les informations enregistrées dans les journaux incluent l'heure de réception de la demande par AWS WAF à partir de votre ressource AWS, des informations détaillées sur la demande et l'action correspondante pour chaque règle demandée. Les exemples de demandes fournissent des détails sur les demandes des trois dernières heures qui correspondaient à l'une de vos règles AWS WAF. Vous pouvez utiliser ces informations pour identifier les signatures de trafic potentiellement malveillantes et créer une nouvelle règle pour refuser ces demandes. Si vous voyez un certain nombre de demandes avec une chaîne de requête aléatoire, veillez à n'autoriser que les paramètres de chaîne de requête pertinents pour le cache de votre application. Cette technique est utile pour atténuer une attaque par destruction du cache contre votre origine.

Si vous êtes abonné à AWS Shield Advanced, vous pouvez faire appel à l'équipe d'intervention AWS Shield (SRT) pour vous aider à créer des règles visant à atténuer une attaque qui nuit à la disponibilité de votre application. Vous pouvez accorder à AWS SRT un accès limité à Shield Advanced et aux API AWS WAF de votre compte. AWS SRT accède à ces API pour placer des mesures d'atténuation sur votre compte uniquement avec votre autorisation explicite. Pour plus d'informations, consultez Support de ce document.

Vous pouvez utiliser AWS Firewall Manager pour configurer et gérer de manière centralisée les règles de sécurité, telles que les protections Shield Advanced et les règles AWS WAF, au sein de votre organisation. Votre compte de gestion AWS Organizations peut désigner un compte administrateur, qui est autorisé à créer des stratégies Firewall Manager. Ces stratégies vous permettent de définir des critères, tels que le type de ressource et les balises, qui déterminent où les règles sont appliquées. Cela est utile lorsque vous avez plusieurs comptes et que vous souhaitez normaliser votre protection.

Pour plus d'informations sur :

- Les règles gérées AWS pour AWS WAF, consultezRègles gérées AWS pourAWS WAF.
- L'utilisation de la restriction géographique pour limiter l'accès à votre distribution CloudFront, consultez Restriction de la distribution géographique de votre contenu.
- En utilisant AWS WAF, consultez
  - Mise en route avec AWS WAF
  - Journalisation des informations de trafic de liste ACL web
  - Affichage d'un exemple de demandes web
- Configuration des règles basées sur les taux, consultez Protection des sites web et des services à l'aide de règles basées sur les taux pour AWS WAF
- Comment gérer le déploiement de règles AWS WAF sur l'ensemble de vos ressources AWS avec Firewall Manager, consultez
  - Premiers pas avec les stratégies de Firewall Manager AWS WAF.
  - Premiers pas avec les stratégies avancées de Firewall Manager Shield.

Réduction de la surface d'attaque

Lors de la création de l'architecture d'une solution AWS, il est également important de limiter les possibilités d'attaque de votre application. Ce concept est connu sous le nom de réduction de surface d'attaque. Les ressources qui ne sont pas exposées à Internet sont plus difficiles à attaquer, ce qui limite les possibilités dont dispose un attaquant pour cibler la disponibilité de votre application.

Par exemple, si vous ne vous attendez pas à ce que les utilisateurs interagissent directement avec certaines ressources, assurez-vous que ces ressources ne sont pas accessibles depuis Internet. De même, n'acceptez pas le trafic provenant d'utilisateurs ou d'applications externes sur des ports ou des protocoles qui ne sont pas nécessaires à la communication.

Dans la section suivante, AWS fournit les bonnes pratiques pour vous aider à réduire votre surface d'attaque et à limiter l'exposition à Internet de votre application.

Rubriques

Dissimuler des ressources AWS (BP1, BP4, BP5)

En règle générale, les utilisateurs peuvent utiliser rapidement et facilement une application sans avoir besoin que les ressources AWS soient entièrement exposées à Internet. Par exemple, lorsque vous avez des instances Amazon EC2 derrière un Elastic Load Balancing, les instances elles-mêmes peuvent ne pas avoir besoin d'être accessibles au public. Au lieu de cela, vous pouvez fournir aux utilisateurs un accès à Elastic Load Balancing sur certains ports TCP et autoriser uniquement Elastic Load Balancing à communiquer avec les instances. Vous pouvez le faire en configurant des groupes de sécurité et des listes de contrôle d'accès (NACL) réseau au sein de votre Amazon Virtual Private Cloud (Amazon VPC). Amazon VPC vous permet de mettre en service une section logiquement isolée du Cloud AWS où vous pouvez lancer des ressources AWS dans un réseau virtuel que vous définissez.

Les groupes de sécurité et les ACL réseau sont similaires, car ils vous permettent de contrôler l'accès aux ressources AWS au sein de votre VPC. Mais les groupes de sécurité vous permettent de contrôler le trafic entrant et sortant au niveau de l'instance, tandis que les ACL réseau offrent des fonctionnalités similaires au niveau du sous-réseau VPC. Il n'y a aucun frais supplémentaire pour l'utilisation de groupes de sécurité ou d'ACL réseau.

Groupes de sécurité et listes de contrôle d'accès réseau (ACL réseau) (BP5)

Vous pouvez choisir de spécifier des groupes de sécurité lorsque vous lancez une instance ou d'associer l'instance à un groupe de sécurité ultérieurement. Tout le trafic vers un groupe de sécurité à partir d'Internet est implicitement refusé, sauf si vous créez une règle autoriser pour autoriser le trafic. Par exemple, si vous avez une application web qui utilise un Elastic Load Balancing et plusieurs instances Amazon EC2, vous pouvez décider de créer un groupe de sécurité pour Elastic Load Balancing (groupe de sécurité Elastic Load Balancing) et un pour les instances (groupe de sécurité du serveur d'applications web). Vous pouvez ensuite créer les règles autoriser pour autoriser le trafic depuis Internet vers le groupe de sécurité ELB et pour autoriser le trafic du groupe de sécurité ELB vers le groupe de sécurité du serveur d'applications web. Cela garantit que le trafic Internet ne peut pas communiquer directement avec vos instances Amazon EC2, ce qui rend plus difficile pour un attaquant d'en savoir plus sur votre application et d'avoir un impact sur celle-ci.

Lorsque vous créez des ACL réseau, vous pouvez spécifier des règles d'autorisation et de refus. Cela est utile si vous souhaitez refuser explicitement certains types de trafic vers votre application. Par exemple, vous pouvez définir des adresses IP (comme plages CIDR), des protocoles et des ports de destination qui doivent être refusés pour l'ensemble du sous-réseau. Si votre site web n'est utilisé que pour le trafic TCP, vous pouvez créer une règle pour interdire tout le trafic UDP, ou inversement. Cette option est utile lorsque vous répondez à des attaques DDoS, car elle vous permet de créer vos propres règles pour atténuer l'attaque lorsque vous connaissez les adresses IP source ou une autre signature.

Si vous êtes abonné à AWS Shield Advanced, vous pouvez enregistrer des adresses IP Elastic en tant que ressources protégées. Les attaques DDoS contre les adresses IP Elastic qui ont été enregistrées en tant que ressources protégées sont détectées plus rapidement, ce qui peut accélérer le délai d'atténuation. Lorsqu'une attaque est détectée, les systèmes d'atténuation DDoS lisent l'ACL réseau qui correspond à l'adresse IP Elastic ciblée et l'appliquent à la frontière du réseau AWS. Cela réduit considérablement le risque d'impact d'un certain nombre d'attaques DDoS au niveau de l'infrastructure.

Pour plus d'informations sur la configuration des groupes de sécurité et des listes de contrôle d'accès réseau afin d'optimiser la résilience DDoS, consultez Comment vous préparer aux attaques DDoS en réduisant votre surface d'attaque.

Pour plus d'informations sur l'utilisation de Shield Advanced avec des adresses IP Elastic en tant que ressources protégées, consultez les étapes pour vous abonner à AWS Shield Advanced.

Protection de votre origine (BP1, BP5)

Si vous utilisez Amazon CloudFront avec une origine située à l'intérieur de votre VPC, vous pouvez vous assurer que seule votre distribution CloudFront peut transférer les demandes vers votre origine. Avec les en-têtes de demande Périphérie-vers-Origine, vous pouvez ajouter ou remplacer la valeur des en-têtes de demande existants lorsque CloudFront transfère les demandes à votre origine. Vous pouvez utiliser les en-têtes personnalisés Origine, par exemple l'en-tête X-Shared-Secret, pour vous aider à vérifier que les demandes adressées à votre origine ont été envoyées depuis CloudFront.

Pour plus d'informations sur la protection de votre origine avec des en-têtes personnalisés Origine, consultez Ajout d'en-têtes personnalisés aux demandes d'origine et Restriction de l'accès aux Application Load Balancers.

Pour obtenir un guide sur la mise en œuvre d'un exemple de solution visant à faire pivoter automatiquement la valeur des en-têtes personnalisés Origine pour la restriction d'accès à l'origine, consultez Comment améliorer la sécurité d'origine Amazon CloudFront avec AWS WAF et Secrets Manager.

Vous pouvez également utiliser une fonction AWS Lambda pour mettre à jour automatiquement les règles de votre groupe de sécurité afin d'autoriser uniquement le trafic CloudFront. Cela améliore la sécurité de votre origine en garantissant que les utilisateurs malveillants ne peuvent pas contourner CloudFront et AWS WAF lorsqu'ils accèdent à votre application web.

Pour plus d'informations sur la façon de protéger votre origine en mettant automatiquement à jour vos groupes de sécurité, consultez l'en-tête X-Shared-Secret, consultez Comment mettre à jour automatiquement vos groupes de sécurité pour Amazon CloudFront et AWS WAF en utilisant AWS Lambda.

Protection des points de terminaison d'API (BP4)

En règle générale, lorsque vous devez exposer une API au public, il existe un risque que l'interface de l'API soit la cible d'une attaque DDoS. Pour réduire les risques, vous pouvez utiliser Amazon API Gateway comme point d'accès aux applications exécutées sur Amazon EC2, AWS Lambda ou ailleurs. En utilisant Amazon API Gateway, vous n'avez pas besoin de vos propres serveurs pour le frontend de l'API et vous pouvez masquer d'autres composants de votre application. En rendant plus difficile la détection des composants de votre application, vous pouvez empêcher ces ressources AWS d'être ciblées par une attaque DDoS.

Lorsque vous utilisez Amazon API Gateway, vous pouvez choisir entre deux types de points de terminaison d'API. La première est l'option par défaut : les points de terminaison d'API optimisés

pour la périphérie auxquels on accède via une distribution Amazon CloudFront. La distribution est toutefois créée et gérée par API Gateway, de sorte que vous n'avez aucun contrôle sur elle. La deuxième option consiste à utiliser un point de terminaison d'API régional accessible à partir de la même région AWS que celle dans laquelle votre API REST est déployée. AWS vous recommande d'utiliser le deuxième type de point de terminaison et de l'associer à votre propre distribution Amazon CloudFront. Cela vous donne le contrôle sur la distribution Amazon CloudFront et la possibilité d'utiliser AWS WAF pour la protection de la couche application. Ce mode vous permet d'accéder à une capacité d'atténuation des attaques DDoS échelonnée sur l'ensemble du réseau périphérique AWS mondial.

Lorsque vous utilisez Amazon CloudFront et AWS WAF avec Amazon API Gateway, configurez les options suivantes :

- Configurez le comportement du cache pour vos distributions afin de transférer tous les en-têtes vers le point de terminaison régional API Gateway. Ce faisant, CloudFront traitera le contenu comme dynamique et évitera la mise en cache du contenu.
- Protégez votre API Gateway contre l'accès direct en configurant la distribution pour inclure l'en-tête personnalisé d'origine `x-api-key`, en définissant la valeur de la clé API dans API Gateway.
- Protégez le backend contre l'excès de trafic en configurant des limites de taux standard ou de rafale pour chaque méthode de vos API REST.

Pour plus d'informations sur la création d'API avec Amazon API Gateway, consultez Mise en route d'Amazon API Gateway.

Techniques opérationnelles

Les techniques d'atténuation présentées dans ce document vous aident à concevoir des applications qui sont intrinsèquement résilientes aux attaques DDoS. Dans de nombreux cas, il est également utile de savoir quand une attaque DDoS cible votre application afin de pouvoir prendre des mesures d'atténuation. Cette section présente les bonnes pratiques pour obtenir une visibilité sur les comportements anormaux, les alertes et l'automatisation, la gestion de la protection à grande échelle et l'engagement d'AWS pour une assistance supplémentaire.

Rubriques

Visibilité

Lorsqu'une mesure opérationnelle clé s'écarte considérablement de la valeur attendue, un attaquant tente peut-être de cibler la disponibilité de votre application. La familiarité avec le comportement normal de votre application signifie que vous pouvez agir plus rapidement lorsque vous détectez une anomalie. Amazon CloudWatch peut vous aider en surveillant les applications que vous exécutez sur AWS. Par exemple, vous pouvez collecter et suivre des métriques, regrouper et contrôler des fichiers journaux, définir des alarmes et réagir automatiquement aux modifications apportées à vos ressources AWS.

Si vous suivez l'architecture de référence résiliente aux attaques DDoS lors de l'architecture de votre application, les attaques courantes de la couche d'infrastructure seront bloquées avant d'atteindre votre application. Si vous êtes abonné à AWS Shield Advanced, vous avez accès à un certain nombre de métriques CloudWatch qui peuvent indiquer que votre application est ciblée. Par exemple, vous pouvez configurer des alarmes pour vous avertir lorsqu'une attaque DDoS est en cours, afin que vous puissiez vérifier l'état de votre application et décider d'engager AWS SRT. Vous pouvez configurer la métrique `DDoSDetected` pour vous informer si une attaque a été détectée. Si vous souhaitez être alerté en fonction du volume d'attaques, vous pouvez également utiliser les métriques `DDoSAttackBitsPerSecond`, `DDoSAttackPacketsPerSecond` ou `DDoSAttackRequestsPerSecond`. Vous pouvez surveiller ces métriques en intégrant CloudWatch à vos propres outils ou en utilisant des outils fournis par des tiers, tels que Slack ou PagerDuty.

Une attaque de la couche applicative peut augmenter de nombreuses métriques Amazon CloudWatch. Si vous utilisez AWS WAF, vous pouvez utiliser CloudWatch pour surveiller et activer des alarmes en cas d'augmentation du nombre de demandes que vous avez définies dans AWS WAF comme étant autorisées, comptabilisées ou bloquées. Cela vous permet de recevoir une notification si le niveau de trafic dépasse ce que votre application peut gérer. Vous pouvez également utiliser Amazon CloudFront, Amazon Route 53, Application Load Balancer, Network Load Balancer, Amazon EC2 et les métriques Auto Scaling qui sont suivies dans CloudWatch pour détecter les changements pouvant indiquer une attaque DDoS.

Le tableau Métriques CloudWatch recommandées répertorie les descriptions des métriques CloudWatch couramment utilisées pour détecter les attaques DDoS et y réagir.

Tableau 3 – Métriques Amazon CloudWatch recommandées

Rubrique	Métrique	Description
AWS Shield Advanced	DDoSDetected	Indique un événement DDoS pour un Amazon Resource Name (ARN).
AWS Shield Advanced	DDoSAttackBitsPerSecond	Nombre d'octets observés pendant un événement DDoS pour un ARN spécifique. Cette métrique est uniquement disponible pour les événements DDoS sur les couches 3/4.
AWS Shield Advanced	DDoSAttackPacketsPerSecond	Nombre de paquets observés pendant un événement DDoS pour un ARN spécifique. Cette métrique est uniquement disponible pour les événements DDoS sur les couches 3/4.
AWS Shield Advanced	DDoSAttackRequestsPerSecond	Nombre de demandes observées lors d'un événement DDoS pour un ARN spécifique. Cette

Rubrique	Métrique	Description
		métrique est uniquement disponible pour les événements DDoS sur la couche 7 et n'est indiquée que pour les événements sur la couche 7 les plus importants.
AWS WAF	AllowedRequests	Nombre de requêtes web autorisées.
AWS WAF	BlockedRequests	Nombre de requêtes web bloquées.
AWS WAF	CountedRequests	Nombre de requêtes web comptabilisées.
AWS WAF	PassedRequests	Nombre de demandes réussies. Ceci est uniquement utilisé pour les demandes qui passent par une évaluation de groupe de règles sans correspondre à aucune des règles du groupe de règles.
Amazon CloudFront	Demandes	Nombre de demandes HTTP/S.
Amazon CloudFront	TotalErrorRate	Pourcentage de toutes les demandes pour lesquelles le code d'état HTTP est 4xx ou 5xx.
Amazon Route 53	HealthCheckStatus	État du point de terminaison de surveillance de l'état.

Rubrique	Métrique	Description
Application Load Balancer	ActiveConnectionCount	Nombre total de connexion s TCP simultanées et actives entre les clients et l'équilibreur de charge et entre l'équilibreur de charge et les cibles.
Application Load Balancer	ConsumedLCUs	Nombre d'unités de capacité d'équilibreur de charge (LCU) utilisées par votre équilibreur de charge.
Application Load Balancer	HTTPCode_ELB_4XX_Count HTTPCode_ELB_5XX_Count	Nombre de codes d'erreur client HTTP 4xx ou 5xx générés par l'équilibreur de charge.
Application Load Balancer	NewConnectionCount	Nombre total de nouvelles connexions TCP établies entre les clients et l'équilibreur de charge et entre l'équilibreur de charge et les cibles.
Application Load Balancer	ProcessedBytes	Nombre total d'octets traités par l'équilibreur de charge.
Application Load Balancer	RejectedConnectionCount	Nombre de connexions rejetées parce que l'équilibreur de charge a atteint le nombre maximal de connexions.
Application Load Balancer	RequestCount	Nombre de demandes qui ont été traitées.

Rubrique	Métrique	Description
Application Load Balancer	TargetConnectionErrorCount	Nombre de connexions qui n'ont pas pu être établies entre l'équilibreur de charge et la cible.
Application Load Balancer	TargetResponseTime	Temps écoulé, en secondes, entre le moment où la demande a quitté l'équilibreur de charge et le moment où la réponse de la cible arrive.
Application Load Balancer	UnHealthyHostCount	Nombre de cibles considérées non saines.
Network Load Balancer	ActiveFlowCount	Nombre total de flux (ou connexions) TCP simultanés provenant des clients vers des cibles.
Network Load Balancer	ConsumedLCUs	Nombre d'unités de capacité d'équilibreur de charge (LCU) utilisées par votre équilibreur de charge.
Network Load Balancer	NewFlowCount	Nombre total de nouveaux flux (ou connexions) TCP établis entre les clients et les cibles pendant la période.
Network Load Balancer	ProcessedBytes	Nombre total d'octets traités par l'équilibreur de charge, y compris les en-têtes TCP/IP.

Rubrique	Métrique	Description
Global Accelerator	NewFlowCount	Nombre total de nouveaux flux (ou connexions) TCP établis entre les clients et les cibles pendant la période.
Global Accelerator	ProcessedBytesIn	Nombre total d'octets entrants traités par l'accélérateur, y compris les en-têtes TCP/IP.
Auto Scaling	GroupMaxSize	Taille maximale du groupe Auto Scaling.
Amazon EC2	CPUUtilization	Pourcentage d'unités de calcul EC2 allouées actuellem ent utilisées dans l'instance.
Amazon EC2	NetworkIn	Nombre d'octets reçus par l'instance sur toutes les interfaces réseau.

Pour plus d'informations sur l'utilisation d'Amazon CloudWatch pour détecter les attaques DDoS sur votre application, consultez Premiers pas avec Amazon CloudWatch.

Pour découvrir un exemple de tableau de bord créé à l'aide de certaines des métriques du tableau précédent, consultez Un système de surveillance de référence personnalisé

AWS inclut plusieurs métriques et alarmes supplémentaires pour vous avertir d'une attaque et pour vous aider à surveiller les ressources de votre application. La console ou l'API AWS Shield fournit un résumé des événements par compte et des détails sur les attaques qui ont été détectées.

En outre, le tableau de bord de l'environnement des menaces mondiales fournit des informations récapitulatives sur toutes les attaques DDoS détectées par AWS. Ces informations peuvent être utiles pour mieux comprendre les menaces DDoS sur un plus grand nombre d'applications, en plus des tendances des attaques et des comparaisons avec les attaques que vous avez pu observer.

Si vous êtes abonné à AWS Shield Advanced, le tableau de bord du service affiche des métriques de détection et d'atténuation supplémentaires, ainsi que des détails sur le trafic réseau pour les événements détectés sur les ressources protégées. AWS Shield évalue le trafic vers votre ressource protégée selon plusieurs dimensions. Lorsqu'une anomalie est détectée, AWS Shield crée un événement et signale la dimension du trafic dans laquelle l'anomalie a été observée. Avec une atténuation placée, cela protège votre ressource contre le trafic excédentaire et le trafic correspondant à une signature d'événement DDoS connue.

Les métriques de détection sont basées sur des flux réseau échantillonnés ou des journaux AWS WAF lorsqu'une ACL web est associée à la ressource protégée. Les métriques d'atténuation sont basées sur le trafic observé par les systèmes d'atténuation des attaques DDoS de Shield. Les métriques d'atténuation sont une mesure plus précise du trafic vers votre ressource.

La métrique des principaux contributeurs du réseau fournit des informations sur la provenance du trafic lors d'un événement détecté. Vous pouvez afficher les contributeurs les plus importants et les trier par aspects tels que le protocole, le port source et les indicateurs TCP. La mesure des principaux contributeurs inclut des métriques pour l'ensemble du trafic observé sur la ressource selon différentes dimensions. Il fournit des dimensions de métrique supplémentaires que vous pouvez utiliser pour comprendre le trafic réseau envoyé à votre ressource lors d'un événement.

Cela comprend également des détails sur les métriques prises automatiquement pour atténuer les attaques DDoS. Ces informations facilitent l'enquête sur les anomalies, l'exploration des dimensions du trafic et une meilleure compréhension des mesures prises par Shield Advanced pour protéger votre disponibilité.

Les journaux de flux VPC constituent un autre outil qui peut vous aider à gagner en visibilité sur le trafic qui cible votre application. Sur un réseau traditionnel, vous pouvez utiliser des journaux de flux réseau pour résoudre les problèmes de connectivité et de sécurité, ainsi que pour vous assurer que les règles d'accès au réseau fonctionnent correctement. En utilisant les journaux de flux VPC, vous pouvez capturer des informations sur le trafic IP qui entre et depuis les interfaces réseau de votre VPC.

Chaque enregistrement du journal de flux inclut les éléments suivants : adresses IP source et destination, ports source et destination, protocole et nombre de paquets et d'octets transférés pendant la fenêtre de capture. Vous pouvez utiliser ces informations pour identifier les anomalies du trafic réseau et identifier un vecteur d'attaque spécifique. Par exemple, la plupart des attaques par réflexion UDP ont des ports source spécifiques, tels que le port source 53 pour la réflexion DNS. Il s'agit d'une signature d'attaque claire que vous pouvez identifier dans l'enregistrement du journal de flux. En réponse, vous pouvez choisir de bloquer le port source spécifique au niveau de l'instance ou

de créer une règle ACL réseau pour bloquer l'ensemble du protocole si votre application n'en a pas besoin.

Pour plus d'informations sur l'utilisation des journaux de flux VPC pour identifier les anomalies réseau et les vecteurs d'attaque DDoS, consultez Journaux de flux VPC et Journaux de flux VPC – Consigner et afficher les flux de trafic réseau.

Gestion de la visibilité et de la protection sur plusieurs comptes

Dans les scénarios où vous opérez sur plusieurs comptes AWS et que vous avez plusieurs composants à protéger, l'utilisation de techniques qui vous permettent d'opérer à grande échelle et de réduire les frais généraux opérationnels augmente vos capacités d'atténuation. Lorsque vous gérez des ressources AWS Shield Advanced protégées dans plusieurs comptes, vous pouvez configurer une surveillance centralisée à l'aide d'AWS Firewall Manager et d'AWS Security Hub. Avec Firewall Manager, vous pouvez créer une stratégie de sécurité qui assure la conformité de la protection DDoS sur tous vos comptes. Vous pouvez utiliser ces deux services ensemble pour gérer vos ressources protégées sur plusieurs comptes et centraliser la surveillance de ces ressources.

Security Hub s'intègre automatiquement à Firewall Manager, ce qui permet aux clients Shield Advanced de consulter les résultats de sécurité dans un tableau de bord unique, ainsi que d'autres alertes de sécurité hautement prioritaires et des statuts de conformité. Par exemple, lorsque Shield Advanced détecte un trafic anormal destiné à une ressource protégée dans n'importe quel compte AWS inclus dans la portée, ce résultat sera visible dans la console Security Hub. S'il est configuré, Firewall Manager peut automatiquement mettre la ressource en conformité en la créant en tant que ressource protégée par Shield Advanced, puis mettre à jour Security Hub lorsque la ressource est dans un état conforme.

Pour plus d'informations sur la surveillance centralisée des ressources protégées par Shield, consultez Configurer la surveillance centralisée des événements DDoS et la correction automatique des ressources non conformes.

Support

Si vous êtes victime d'une attaque, vous pouvez également bénéficier d'une assistance AWS pour évaluer la menace et examiner l'architecture de votre application, ou vous pouvez demander une autre assistance. Il est important de créer un plan de réponse aux attaques DDoS avant qu'un

événement ne se produise. Les bonnes pratiques décrites dans ce document sont destinées à être des mesures proactives que vous mettez en œuvre avant de lancer une application, mais des attaques DDoS contre votre application peuvent toujours se produire. Passez en revue les options de cette section pour déterminer les ressources de support les mieux adaptées à votre scénario. L'équipe chargée de votre compte peut évaluer votre cas d'utilisation et votre application, et répondre à vos questions ou difficultés spécifiques.

Si vous exécutez des charges de travail de production sur AWS, pensez à vous abonner à Business Support, qui vous fournit un accès 24 heures sur 24, 7 jours sur 7, aux ingénieurs du support cloud qui peuvent vous aider à résoudre les problèmes d'attaques DDoS. Si vous exécutez des charges de travail critiques, pensez à l'assistance aux entreprises, qui permet d'ouvrir des cas critiques et de recevoir la réponse la plus rapide d'un ingénieur de support cloud senior.

Si vous êtes abonné à AWS Shield Advanced et que vous êtes également abonné à Business Support ou Enterprise Support, vous pouvez configurer l'engagement proactif Shield. Il vous permet de configurer des surveillances de l'état, de vous associer à vos ressources et de fournir des informations de contact pour les opérations 24 heures sur 24 Lorsque Shield détecte des signes de DDoS et que la surveillance de l'état de vos applications montre des signes de dégradation, AWS SRT vous contacte de manière proactive. Il s'agit de notre modèle d'engagement recommandé, car il permet les temps de réponse AWS SRT les plus rapides et permet à AWS SRT de commencer le dépannage avant même que le contact n'ait été établi avec vous.

La fonctionnalité d'engagement proactif nécessite que vous configuriez une surveillance de l'état Route 53 qui mesure avec précision l'état de votre application et qui est associée à la ressource protégée par Shield Advanced. Une fois qu'une surveillance de l'état Route 53 est associée dans la console Shield, le système de détection Shield Advanced utilise la surveillance de l'état comme indicateur de l'état de votre application. La fonction de détection basée sur l'état de Shield Advanced garantit que vous êtes averti et que des mesures d'atténuation sont mises en place plus rapidement lorsque votre application est défectueuse. AWS SRT vous contactera pour déterminer si l'application défectueuse est ciblée par une attaque DDoS et mettra en place des mesures d'atténuation supplémentaires si nécessaire.

L'achèvement de la configuration de l'engagement proactif inclut l'ajout de détails de contact dans la console Shield. AWS SRT utilisera ces informations pour vous contacter. Vous pouvez configurer jusqu'à 10 contacts et fournir des notes supplémentaires si vous avez des exigences ou des préférences de contact spécifiques. Les contacts d'engagement proactifs doivent occuper un rôle 24 heures sur 24, 7 jours sur 7, comme un centre des opérations de sécurité ou une personne immédiatement disponible.

Vous pouvez activer l'engagement proactif pour toutes les ressources ou pour certaines ressources de production clés pour lesquelles le temps de réponse est essentiel. Pour ce faire, affectez des contrôles de santé uniquement à ces ressources.

Vous pouvez également passer à AWS SRT en créant un cas AWS Support à l'aide de la console AWS Support ou de l'API Support si vous avez un événement lié à un DDoS qui affecte la disponibilité de votre application.

Conclusion

Les bonnes pratiques présentées dans ce document peuvent vous aider à créer une architecture résiliente aux attaques DDoS qui protège la disponibilité de votre application en empêchant de nombreuses attaques DDoS courantes au niveau de l'infrastructure et de la couche application. La mesure dans laquelle vous suivez ces bonnes pratiques lorsque vous concevez votre application aura une influence sur le type, le vecteur et le volume des attaques DDoS que vous pouvez atténuer. Vous pouvez intégrer la résilience sans vous abonner à un service d'atténuation des attaques DDoS. En choisissant de vous abonner à AWS Shield Advanced, vous bénéficiez de fonctions supplémentaires de support, de visibilité, d'atténuation et de protection des coûts qui protègent davantage une architecture d'application déjà résiliente.

Participants

Ont contribué à la préparation du présent document :

- Jeffrey Lyon, Protection du périmètre AWS
- Rodrigo Ferroni, Spécialiste de la sécurité AWS TAM
- Dmitriy Novikov, Architecte de solutions AWS
- Achraf Souk, Architecte de solutions AWS
- Yoshihisa Nakatani, Architecte de solutions AWS

Ressources

Suggestions de lecture :

- Bonnes pratiques pour la réduction des attaques DDoS sur AWS
- Directives de mise en œuvre d'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! Établir des bases solides avec Amazon CloudFront, AWS Shield et 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 : Protection hautes performances contre les DDoS avec AWS

Révisions du document

Pour être informé des mises à jour de ce livre blanc, abonnez-vous au flux RSS.

update-history-change	update-history-description	update-history-date
Livre blanc mis à jour	Mise à jour pour inclure les dernières recommandations et fonctionnalités. AWS Global Accelerator est ajouté dans le cadre d'une protectio n complète en périphérie. AWS Firewall Manager pour la surveillance centralisée des événements DDoS et la correction automatique des ressources non conformes.	21 septembre 2021
Livre blanc mis à jour	Mise à jour pour clarifier le blocage du cache dans la section Détecter et filtrer les requêtes web malveillantes (BP1, BP2) et l'utilisation de l'ELB et de l'ALB dans la section Mettre à l'échelle pour absorber (BP6). Diagramme s mis à jour et tableau 2, marqué « Choix de la région » comme BP8. Section BP7 mise à jour avec plus de détails.	18 décembre 2019
Livre blanc mis à jour	Mise à jour pour inclure la journalisation AWS WAF en tant que bonne pratique.	1er décembre 2018

Livre blanc mis à jour	Mise à jour pour inclure AWS Shield, les fonctions AWS WAF, AWS Firewall Manager et les bonnes pratiques associées.	1er juin 2018
Livre blanc mis à jour	Ajout de conseils d'architecture prescriptifs et mise à jour pour inclure AWS WAF.	1er juin 2016
Publication initiale	Livre blanc publié.	1er juin 2015

Mentions légales

Les clients sont chargés d'évaluer par eux-même les informations contenues dans ce document. Le présent document : (a) est fourni à titre informatif uniquement, (b) représente les offres et pratiques actuelles de produits AWS, qui sont susceptibles d'être modifiées sans préavis, et (c) ne crée aucun engagement ou assurance de la part d'AWS et de ses affiliés, fournisseurs ou concédants de licences. Les produits ou services AWS sont fournis « en l'état » sans garantie, représentation ou condition, de quelque nature que ce soit, explicite ou implicite. Les responsabilités et obligations d'AWS envers ses clients sont déterminées par les contrats AWS, et le présent document ne fait pas partie d'un contrat entre AWS et ses clients, ni le modifie.