Meilleures pratiques de l'AWS pour la résilience des DDoS

Juin 2018

Avis

Ce document est fourni à titre d'information uniquement. Il représente les offres de produits et les pratiques actuelles de l'AWS à la date de publication de ce document, qui sont susceptibles d'être modifiées sans préavis. Les clients sont responsables de leur propre évaluation indépendante des informations contenues dans ce document et de toute utilisation des produits ou services d'AWS, chacun étant fourni "tel quel" sans garantie d'aucune sorte, qu'elle soit expresse ou implicite. Ce document ne crée aucune garantie, représentation, engagement contractuel, condition ou assurance de la part d'AWS, de ses affiliés, fournisseurs ou concédants de licence. Les responsabilités
et les responsabilités de l'AWS envers ses clients sont contrôlées par les accords de l'AWS, et le présent document ne fait pas partie d'un accord entre l'AWS et ses clients, ni ne le modifie.

Contenu

Résumé

Ce document est destiné aux clients qui souhaitent améliorer la résilience de leurs applications fonctionnant sur Amazon Web Services (AWS) contre les attaques par déni de service distribué (DDoS). Il donne un aperçu des attaques DDoS, des capacités offertes par les AWS, des techniques d'atténuation et d'une architecture de référence résistante aux DDoS qui peut être utilisée comme guide pour aider à protéger la disponibilité des applications.

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 de l'AWS. Chaque section comporte des liens vers la documentation de l'AWS qui fournit plus de détails sur les meilleures pratiques ou capacités.

Introduction

Vous vous efforcez de protéger votre entreprise contre l'impact des attaques par déni de service distribué (DDoS), ainsi que d'autres cyberattaques. Vous voulez conserver la confiance de vos clients dans votre service en maintenant la disponibilité et la réactivité de votre application. Et vous voulez éviter des coûts directs inutiles lorsque votre infrastructure doit s'adapter en réponse à une attaque.

AWS s'engage à vous fournir des outils, des meilleures pratiques et des services pour vous aider à assurer une haute disponibilité, une sécurité et une résilience pour vous défendre contre les mauvais acteurs de l'internet.

Dans ce livre blanc, nous vous donnons des conseils prescriptifs sur les DDoS. Nous décrivons différents types d'attaques, telles que les attaques de la couche infrastructure et les attaques de la couche application, et nous expliquons quelles sont les meilleures pratiques pour gérer chaque type d'attaque. Nous décrivons également les services et les fonctionnalités qui s'inscrivent dans une stratégie d'atténuation des DDoS, et comment chacun d'entre eux peut être utilisé pour aider à protéger vos applications.

Attaques par déni de service

Une attaque par déni de service (DDS) est une tentative délibérée de rendre votre site web ou votre application indisponible pour les utilisateurs, par exemple en l'inondant de trafic réseau. Pour y parvenir, les attaquants utilisent diverses techniques qui consomment une grande quantité de bande passante ou qui bloquent d'autres ressources du système, perturbant ainsi l'accès des utilisateurs légitimes. Dans sa forme la plus simple, un attaquant isolé utilise une source unique pour exécuter une attaque DoS contre une cible, comme le montre le schéma suivant (figure 1).

Figure 1 : Diagramme de l'attaque DOS

Mais dans une attaque par déni de service distribué (DDoS), un attaquant utilise plusieurs sources - qui peuvent être des groupes distribués d'ordinateurs, de routeurs, de dispositifs IoT et d'autres terminaux infectés par des logiciels malveillants - pour orchestrer une attaque contre une cible. Comme l'illustre le diagramme d'attaque DDoS suivant (figure 2), un réseau d'hôtes compromis participe à l'attaque, générant un flot de paquets ou de demandes pour submerger la cible.

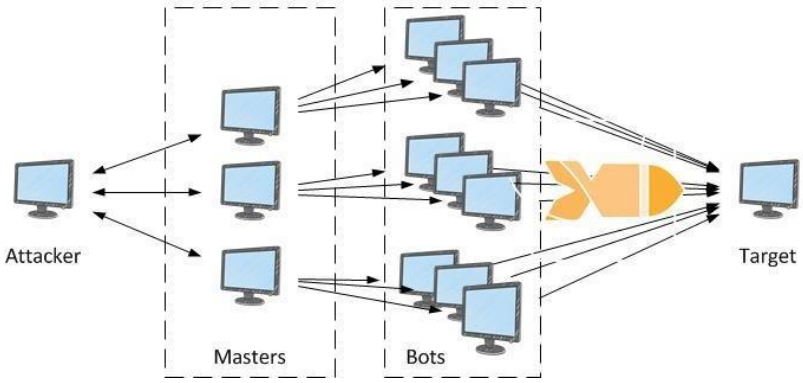

Figure 2 : Diagramme d'une attaque DDoS

Les attaques DDoS sont les plus courantes aux couches 3, 4, 6 et 7 du modèle d'interconnexion des systèmes ouverts (OSI), qui est décrit dans le tableau suivant (tableau 1). Les attaques des couches 3 et 4 correspondent aux couches réseau et transport du modèle OSI. Nous les appellerons collectivement les attaques de la *couche infrastructure*. Les attaques des couches 6 et 7 correspondent aux couches "Présentation" et "Transport" du modèle OSI.

Couches d'application du modèle OSI. Nous les aborderons ensemble en tant qu'*attaques des couches d'application*. Nous décrivons des exemples de ces types d'attaques dans les sections suivantes.

#	Couche	Unité	Description	Exemples de vecteurs
7	Candidature	Données	Processus de candidature en réseau	Inondations HTTP, inondations d'interrogation DNS
6	Présentation	Données	Représentation et cryptage des données	Abus de TLS
5	Session	Données	Communication entre hôtes	N/A
4	Transport	Segments	Connexions de bout en bout et fiabilité	Inondations du SYN
3	Réseau	Paquets	Détermination du chemin et adressage logique	Les attaques de réflexion de l'UDP
2	Liaison de données	Cadres	Adressage physique	N/A
1	Physique	Bits	Médias, signaux et transmission binaire	N/A

Tableau 1 : Modèle d'interconnexion des systèmes ouverts (OSI).

Attaques de la couche d'infrastructure

Les attaques DDoS les plus courantes, les attaques par réflexion UDP et les inondations SYN, sont des attaques de la couche infrastructurelle. Un attaquant peut utiliser l'une ou l'autre de ces méthodes pour générer de gros volumes de trafic qui peuvent inonder la capacité d'un réseau ou immobiliser des ressources sur des systèmes tels qu'un serveur, un pare-feu, un 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 leur capacité plus rapidement que l'inondation du trafic entrant. Cette capacité supplémentaire sert à filtrer ou à absorber le trafic d'attaque, ce qui permet à votre système et à votre application de répondre au trafic de vos clients légitimes.

Attaques de réflexion de l'UDP

Les attaques par réflexion UDP exploitent le fait que l'UDP est un protocole apatride. Les attaquants peuvent créer un paquet de requête UDP valide qui indique l'adresse IP de la cible de l'attaque comme étant l'adresse IP source de l'UDP. L'attaquant a alors falsifié, ou "spoofé", l'adresse IP source du paquet de requête UDP. Un attaquant envoie alors le paquet UDP contenant l'adresse IP source usurpée à un serveur intermédiaire. Le serveur est amené par ruse à envoyer ses paquets de réponse UDP à l'adresse IP victime ciblée plutôt que de les renvoyer à l'adresse

l'adresse IP de l'attaquant. Le serveur intermédiaire est utilisé parce qu'il génère une réponse plusieurs fois plus importante que le paquet de requête, ce qui amplifie effectivement la quantité de trafic d'attaque envoyé à l'adresse IP cible.

Le facteur d'amplification, qui est le rapport entre la taille de la réponse et la taille de la demande, varie en fonction du protocole utilisé par l'attaquant : DNS, NTP ou SSDP. Par exemple, le facteur d'amplification pour le DNS peut être de 28 à 54 fois le nombre d'octets d'origine. Ainsi, si un attaquant envoie une demande de 64 octets à un serveur DNS, il peut générer plus de 3 400 octets de trafic indésirable vers une cible d'attaque. La tactique de réflexion et l'effet d'amplification sont illustrés dans le diagramme suivant (figure 3).

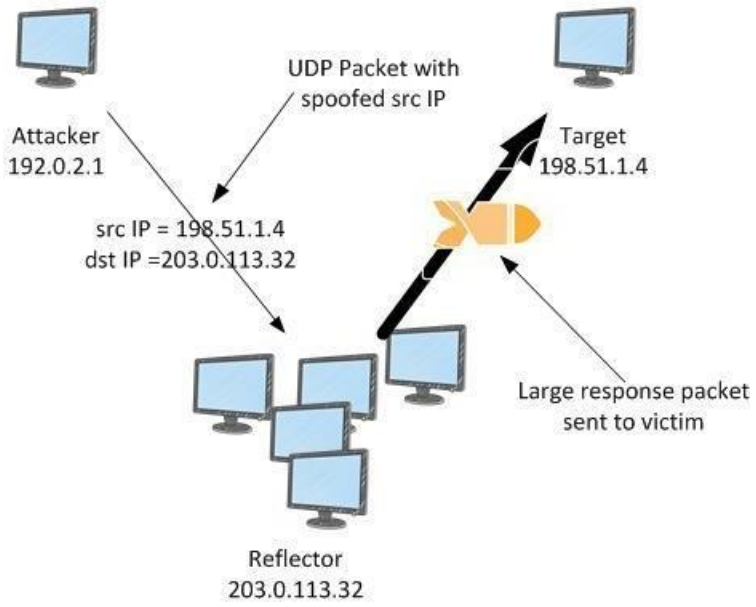

Figure 3 : Attaque par réflexion de l'UDP

Attaques des inondations du SYN

Lorsqu'un utilisateur se connecte à un service TCP, comme un serveur web, son client envoie un paquet SYN (synchronisation). Le serveur renvoie un paquet SYN-ACK en guise d'accusé de réception et, finalement, le client répond avec un paquet ACK, ce qui complète la poignée de main à trois voies attendue. Cette poignée de main typique est illustrée dans le schéma suivant (figure 4) :

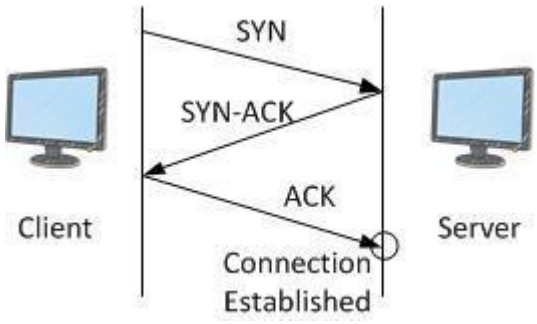

Figure 4 : SYN poignée de main à trois voies

Lors d'une attaque par flood SYN, un client malveillant envoie un grand nombre de paquets SYN, mais n'envoie jamais les derniers paquets ACK pour compléter les poignées de main. Le serveur est laissé en attente d'une réponse aux connexions TCP à moitié ouvertes et finit par manquer de capacité pour accepter de nouvelles connexions TCP. Cela peut empêcher de nouveaux utilisateurs de se connecter au serveur. Les inondations SYN peuvent atteindre jusqu'à 100s de Gbps, mais l'attaque ne porte pas sur le volume de trafic SYN mais plutôt sur le blocage des connexions serveur disponibles, ce qui entraîne l'absence de ressources pour les connexions légitimes.

Attaques de la couche d'application

Un attaquant peut cibler l'application elle-même en utilisant une attaque de la couche 7 ou de la couche application. Dans ces attaques, similaires aux attaques de l'infrastructure d'inondation SYN, l'attaquant tente de surcharger des fonctions spécifiques d'une application afin de rendre l'application indisponible ou extrêmement peu réactive pour les utilisateurs légitimes. Parfois, cela peut être réalisé avec de très faibles volumes de requêtes 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 applicative, on peut citer les inondations HTTP, les attaques de destruction de cache et les inondations XML-RPC WordPress.

Dans une **attaque par flood HTTP**, un attaquant envoie des requêtes HTTP qui semblent provenir d'un utilisateur réel de l'application web. Certaines inondations HTTP ciblent une ressource spécifique, tandis que les inondations HTTP plus complexes tentent d'émuler l'interaction humaine avec l'application. Cela peut accroître la difficulté d'utiliser des techniques d'atténuation courantes comme la limitation du débit des requêtes.

Les **attaques de 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 distribution de contenu (CDN). Au lieu de pouvoir renvoyer les résultats mis en cache, le CDN doit contacter le serveur d'origine pour chaque requête de page, et ces récupérations d'origine provoquent une tension supplémentaire sur le serveur web d'application.

Lors d'une **attaque par flood XML-RPC de WordPress**, également appelée pingback flood de WordPress, un attaquant utilise abusivement la fonction API XML-RPC d'un site web hébergé sur le logiciel de gestion de contenu WordPress pour générer un flot de requêtes HTTP. La fonction pingback permet à un site web hébergé sur WordPress (Site A) de notifier à un autre site WordPress (Site B) que le Site A a créé un lien vers le Site B. Le Site B tente alors de récupérer le Site A pour vérifier l'existence du lien. Dans une inondation de pingback, l'attaquant utilise abusivement cette capacité pour amener le site B à attaquer le site A. Ce type d'attaque a une signature claire : "WordPress" est généralement présent dans le "User-Agent" de l'en-tête de la requête HTTP.

Il existe également d'autres formes de trafic malveillant qui peuvent avoir un impact sur la disponibilité d'une application. Les scraper bots automatisent les tentatives d'accès à une application web pour voler du contenu ou enregistrer des informations sur la concurrence, comme les prix. Les attaques par force brute et par bourrage de références sont des efforts programmés pour obtenir un accès non autorisé aux zones sécurisées d'une application. Il ne s'agit pas strictement d'attaques DDoS, mais leur nature automatisée peut ressembler à une attaque DDoS et elles peuvent être atténuées par la mise en œuvre de certaines des meilleures pratiques qui seront abordées plus loin dans ce document.

Les attaques de la couche application peuvent également viser les services du système de noms de domaine (DNS). La plus courante de ces attaques est une inondation de requêtes DNS dans 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 de destruction de cache dans lequel l'attaquant randomise la chaîne de sous-domaines pour contourner le cache DNS local de n'importe quel résolveur donné. En conséquence, le résolveur ne peut pas tirer profit des requêtes de domaine mises en cache et doit au contraire contacter de manière répétée le serveur DNS faisant autorité, ce qui amplifie l'attaque.

Si une application web est livrée via TLS, un attaquant peut également choisir d'attaquer le processus de négociation TLS. Le TLS est coûteux en termes de calcul, de sorte qu'un attaquant peut réduire la disponibilité d'un serveur en envoyant des données inintelligibles. Dans une variante de cette attaque, un attaquant complète la poignée de main TLS mais renégocie perpétuellement la méthode de cryptage. Ou bien un attaquant peut tenter d'épuiser les ressources du serveur en ouvrant et en fermant de nombreuses sessions TLS.

Techniques d'atténuation

Certaines formes d'atténuation des DDoS sont automatiquement incluses dans les services AWS. Vous pouvez encore améliorer votre résistance aux DDoS en utilisant une architecture AWS avec des services spécifiques et en mettant en œuvre des meilleures pratiques supplémentaires.

Tous les clients AWS bénéficient des protections automatiques de la norme AWS Shield, sans frais supplémentaires. AWS Shield Standard se défend contre les attaques DDoS les plus courantes et les plus fréquentes des couches réseau et transport qui ciblent votre site web ou vos applications. Cette protection est offerte sur tous les services AWS et dans chaque région AWS, sans frais supplémentaires. Dans les régions AWS, les attaques DDoS sont détectées par un système qui établit automatiquement une base de trafic, identifie les anomalies et, si nécessaire, crée des mesures d'atténuation. Ce système d'atténuation offre une protection contre de nombreuses attaques courantes de la couche d'infrastructure. Vous pouvez utiliser la norme AWS Shield dans le cadre d'une architecture résistante aux attaques DDoS pour protéger les applications web et non web.

En outre, vous pouvez tirer parti des services AWS qui fonctionnent à partir de sites périphériques, comme Amazon CloudFront et Amazon Route 53, pour mettre en place une protection de disponibilité complète contre toutes les attaques connues de la couche d'infrastructure. L'utilisation de ces services, qui font partie du réseau AWS Global Edge Network, peut améliorer la résilience DDoS de votre application lorsque vous servez le trafic d'applications web à partir de sites périphériques répartis dans le monde entier.

L'utilisation d'Amazon CloudFront et d'Amazon Route 53 présente plusieurs avantages spécifiques :

- Les systèmes d'atténuation des DDoS AWS Shield qui sont intégrés aux services de pointe AWS, réduisant le temps d'atténuation de quelques minutes à une sous-seconde.
- Stateless SYN Techniques d'atténuation des inondations qui utilisent des proxy et vérifient les connexions entrantes avant de les transmettre au service protégé.
- Des systèmes automatiques d'ingénierie du trafic qui peuvent disperser ou isoler l'impact des attaques DDoS volumétriques de grande envergure.
- La défense de la couche applicative lorsqu'elle est combinée avec le WAF AWS qui ne nécessite pas de modifier votre architecture applicative actuelle (par exemple dans une région AWS ou un centre de données sur site

Le transfert de données entrantes sur l'AWS est gratuit et vous ne payez pas pour le trafic d'attaque DDoS qui est atténué par le bouclier AWS.

Pour une architecture de référence résistante aux DDoS qui inclut les services du réseau AWS Global Edge Network, voir le schéma suivant (figure 5).

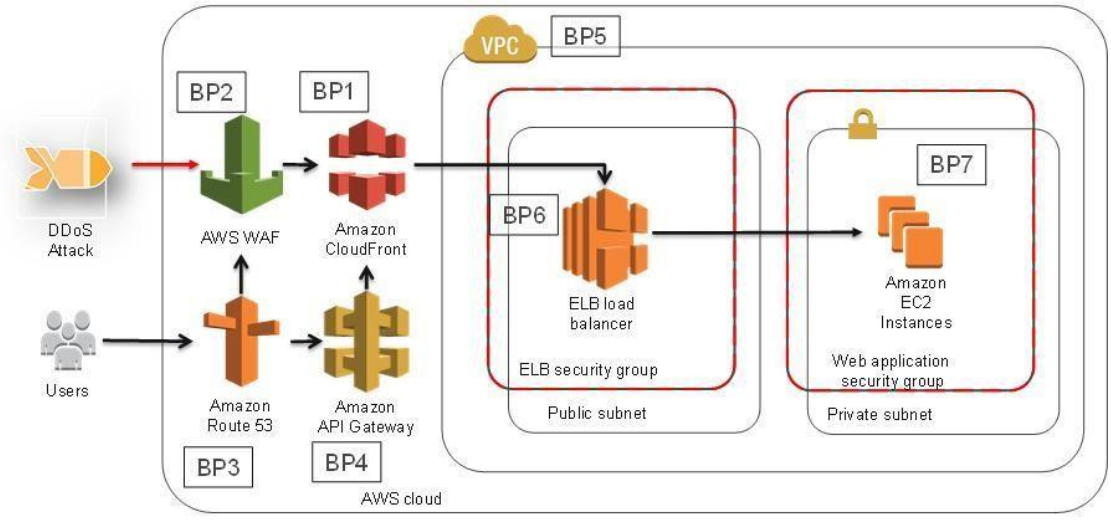

Figure 5 : Architecture de référence résistante aux DDoS.

Cette architecture de référence comprend plusieurs services AWS qui peuvent vous aider à améliorer la résilience de votre application web contre les attaques DDoS. Pour un résumé de ces services et des capacités qu'ils peuvent fournir, voir le tableau suivant (tableau 2). Nous avons marqué chaque service d'un "indicateur de meilleure pratique" (BP1, BP2, etc.) afin que vous puissiez facilement vous y référer ici lorsque vous lirez le reste de ce document. Par exemple, une section à venir traite des capacités fournies par Amazon CloudFront et inclut l'indicateur de meilleures pratiques BP1.

	Emplacements des AWS Edge		**Régions AWS**			
	Amazon CloudFront (BP1) avec AWS WAF (BP2)	Amazon Route 53 (BP3)	Élastique Charger Équilibre (BP6)	Amazon API Passerelle (BP4)	Amazon VPC (BP5)	Amazone EC2 avec Auto Mise à l'échelle (BP7)
Atténuation des attaques de la couche 3 (par exemple, réflexion UDP)	✔	✔	✔	✔	✔	✔

	Emplacements des AWS Edge		Régions AWS			
Atténuation des attaques de la couche 4 (par exemple, inondation SYN)	✔	✔	✔	✔		
Atténuation des attaques de la couche 6 (par exemple, TLS)	✔		✔	✔		
Réduire la surface d'attaque	✔	✔	✔	✔	✔	
Échelle permettant d'absorber le trafic de la couche d'application	✔	✔	✔	✔		✔
Atténuation des attaques de la couche 7 (couche d'applicatio n)	✔	✔	✔ (si utilisé avec AWS WAF)			
Isolement géographique et dispersion du trafic excédentaire, et attaques DDoS plus importantes	✔	✔				

Tableau 2 : Résumé des meilleures pratiques.

Un autre moyen d'améliorer votre capacité de réaction et d'atténuation des attaques DDoS est de vous abonner à AWS Shield Advanced. Ce service optionnel d'atténuation des attaques DDoS vous aide à protéger une application hébergée dans une région AWS ou en dehors de l'AWS. Le service est disponible dans le monde entier pour Amazon CloudFront et Amazon Route 53. Il est également disponible dans certaines régions AWS pour les équilibres de charge classiques (CLB), les équilibres de charge applicatifs (ALB) et les adresses IP élastiques (EIP). L'utilisation de AWS Shield Advanced avec les EIP vous permet de protéger les instances Network Load Balancer (NLB) ou Amazon EC2.

Avec AWS Shield Advanced, vous bénéficiez des avantages supplémentaires suivants :

- Accès à l'AWS DDoS Response Team (DRT) pour l'aide à l'atténuation des attaques DDoS qui ont un impact sur la disponibilité des applications.

- La visibilité des attaques DDoS en utilisant la console de gestion AWS, l'API et les mesures et alarmes d'Amazon CloudWatch.
- Accès au tableau de bord "Global Threat Environment", qui donne un aperçu des attaques DDoS observées et atténuées par l'AWS.
- Accès à AWS WAF, sans frais supplémentaires, pour l'atténuation des attaques DDoS de la couche application (lorsqu'il est utilisé avec Amazon CloudFront ou ALB).
- Baselining automatique des attributs du trafic web, lorsqu'il est utilisé avec AWS WAF.
- Accès au gestionnaire de pare-feu AWS, sans frais supplémentaires, pour une application automatisée des politiques. Ce service permet aux administrateurs de la sécurité de contrôler et de gérer de manière centralisée les règles de l'AWS WAF.
- Des seuils de détection sensibles qui acheminent le trafic plus tôt dans le système d'atténuation des DDoS et peuvent améliorer le temps d'atténuation des attaques contre Amazon EC2 ou NLB, lorsqu'ils sont utilisés avec l'EIP.
- Protection des coûts qui vous permet de demander un remboursement limité des coûts liés au scalage qui résultent d'une attaque DDoS.
- Enhanced Service Level Agreement qui est spécifique aux clients d'AWS Shield Advanced.

Pour une liste complète des fonctionnalités avancées de AWS Shield et pour en savoir plus sur AWS Shield, voir AWS Shield - Managed DDoS Protection.

Dans les sections suivantes, nous décrirons plus en détail chacune des meilleures pratiques recommandées pour la réduction des DDoS. Pour un guide rapide et facile à mettre en œuvre sur la construction d'une couche de réduction des DDoS pour les applications web statiques ou dynamiques, voir Comment aider à protéger les applications web dynamiques contre les attaques DDoS en utilisant Amazon CloudFront et Amazon Route 53.

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

Dans un environnement de centre de données traditionnel, vous pouvez atténuer les attaques DDoS de la couche infrastructurelle en utilisant des techniques telles que le surapprovisionnement en capacité, le déploiement de systèmes d'atténuation des DDoS ou l'épuration du trafic à l'aide de services d'atténuation des DDoS. Sur l'AWS, les capacités d'atténuation des DDoS sont automatiquement fournies, mais vous pouvez optimiser la résilience de votre application aux DDoS en choisissant l'architecture qui exploite au mieux ces capacités et qui vous permet également de faire face à l'excès de trafic.

Pour atténuer les attaques DDoS volumétriques, il est essentiel de veiller à ce que la capacité et la diversité de transit soient suffisantes et de protéger vos ressources AWS, comme les instances Amazon EC2, contre les attaques de trafic.

Taille de l'instance (BP7)

Amazon EC2 offre une capacité de calcul redimensionnable qui vous permet d'augmenter ou de réduire rapidement votre capacité de calcul en fonction de l'évolution de vos besoins. Vous pouvez changer d'échelle horizontalement en ajoutant des instances à votre application, et verticalement en utilisant des instances plus grandes. Certains types d'instances d'Amazon EC2 prennent en charge des fonctionnalités qui peuvent plus facilement gérer de gros volumes de trafic, par exemple les interfaces réseau 25 Gigabit et Enhanced Networking.

Avec des interfaces réseau de 25 gigabits, chaque instance peut supporter un plus grand volume de trafic. Cela permet d'éviter l'encombrement des interfaces pour le trafic qui a atteint l'instance Amazon EC2.
Les instances qui prennent en charge l'Enhanced Networking offrent des performances d'entrée/sortie plus élevées et une utilisation plus faible du processeur par rapport aux implémentations traditionnelles. Cela améliore la capacité de l'instance à gérer le trafic avec des volumes de paquets plus importants.

La fonction 25 Gigabit est disponible sur les plus grandes tailles d'instances, par exemple, M4. Pour en savoir plus sur les instances Amazon EC2 qui prennent en charge les interfaces réseau 25 Gigabit et Enhanced Networking, voir Types d'instances Amazon EC2. Pour savoir comment activer l'Enhanced Networking, voir Enabling Enhanced Networking on Linux Instancesin a VPC.

Choix de la région (BP7)

Les services AWS sont disponibles dans de nombreux endroits du monde. Ces zones géographiquement séparées sont appelées des régions. Lors de l'élaboration de votre application, vous pouvez choisir une ou plusieurs régions en fonction de vos besoins. Les considérations communes comprennent la performance, le coût et la souveraineté des données. Dans chaque région, l'AWS donne accès à un ensemble unique de connexions Internet et de relations de peering afin d'offrir une latence et un débit optimaux aux utilisateurs de ces régions.

Il est également important de tenir compte de la résilience des DDoS lorsque vous choisissez les régions pour votre candidature. De nombreuses régions sont proches d'échanges internet et disposent donc d'une plus grande connectivité aux principaux réseaux. La proximité d'échanges où les opérateurs internationaux et les grands homologues sont fortement présents peut vous donner la capacité internet nécessaire pour atténuer les attaques de grande envergure.

Pour en savoir plus sur le choix d'une région, voir Régions et zones de disponibilité. Vous pouvez également demander à votre équipe de compte les caractéristiques de chaque région que vous envisagez, afin de vous aider à prendre une décision éclairée.

Équilibrage des charges (BP6)

Les attaques DDoS de grande envergure peuvent dépasser la capacité d'une seule instance d'Amazon EC2, aussi l'ajout de l'équilibrage de charge peut aider votre résilience. Vous pouvez choisir parmi plusieurs options pour atténuer une attaque en équilibrant la charge du trafic excédentaire. Avec l'équilibrage de charge élastique (ELB), vous pouvez réduire le risque de surcharge de votre application en répartissant le trafic sur de nombreuses instances de backend. L'ELB peut s'adapter 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 construites dans Amazon VPC, il existe deux types d'ELB à prendre en compte, selon le type d'application : Application Load Balancer (ALB) ou Network Load Balancer (NLB).

Pour les applications web, vous pouvez utiliser l'ALB pour acheminer le trafic en fonction de son contenu et n'accepter que les demandes web bien formées. Cela signifie que de nombreuses attaques DDoS courantes, comme les inondations SYN ou les attaques par réflexion UDP, seront bloquées par l'ALB, protégeant ainsi votre application contre l'attaque. Lorsque l'ALB détecte ces types d'attaques, il se met automatiquement à l'échelle pour absorber le trafic supplémentaire. Pour en savoir plus sur la protection des applications web avec l'ALB, voir la section Démarrage avec les équilibreurs de charge applicatifs.

Pour les applications basées sur le protocole TCP, vous pouvez utiliser le NLB pour acheminer le trafic vers les instances Amazon EC2 avec une latence ultra-faible. Lorsque vous créez un NLB, une interface réseau est créée pour chaque zone de disponibilité (AZ) que vous activez. Vous avez la possibilité d'attribuer une adresse IP élastique (EIP) par sous-réseau activé pour l'équilibreur de charge. L'un des points clés de la NLB est que tout trafic qui atteint l'équilibreur de charge sur un auditeur valide sera acheminé vers vos instances Amazon EC2, et non pas absorbé. Pour en savoir plus sur la protection des applications TCP avec NLB, consultez la section Démarrage avec les répartiteurs de charge réseau.

Livrer à l'échelle en utilisant les emplacements AWS Edge (BP1, BP3)

L'accès à des connexions Internet diverses et à grande échelle peut augmenter considérablement 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 sites AWS en périphérie fournissent une couche supplémentaire d'infrastructure réseau qui offre ces avantages à toute application web qui utilise Amazon CloudFront et Amazon Route 53. Lorsque vous utilisez ces services, votre contenu est servi et les requêtes DNS sont résolues à partir d'emplacements qui sont souvent plus proches de vos utilisateurs.

Application Web Delivery at the Edge (BP1)

Amazon CloudFront est un service qui peut être utilisé pour diffuser l'intégralité de votre site web, y compris les contenus statiques, dynamiques, en streaming et interactifs. Des connexions TCP persistantes et des paramètres de temps de vie variables (TTL) peuvent être utilisés pour décharger le trafic de votre site d'origine, même si vous n'êtes pas servir du contenu cachable. Ces caractéristiques signifient que l'utilisation d'Amazon CloudFront réduit le nombre de requêtes et de connexions TCP vers votre origine, ce qui contribue à protéger votre application web contre les inondations HTTP. Amazon CloudFront n'accepte que les connexions bien formées, ce qui permet d'éviter que de nombreuses attaques DDoS courantes, comme les inondations SYN et les attaques par réflexion UDP, n'atteignent votre origine. Les attaques DDoS sont également isolées géographiquement près de la source, ce qui empêche le trafic d'avoir un impact sur d'autres endroits. Ces fonctionnalités peuvent améliorer considérablement votre capacité à continuer à servir le trafic aux utilisateurs lors d'attaques DDoS de grande envergure. Vous pouvez utiliser Amazon CloudFront pour protéger une origine sur AWS ou ailleurs sur Internet.

Si vous utilisez Amazon S3 pour diffuser du contenu statique sur Internet, vous devez utiliser Amazon CloudFront pour protéger votre seau. Vous pouvez utiliser la fonction OAI (Origin Access Identify) pour vous assurer que les utilisateurs n'accèdent à vos objets qu'en utilisant les URL de CloudFront. Pour en savoir plus sur l'OAI, voir Utiliser une identité d'accès à l'origine pour restreindre l'accès à votre contenu Amazon S3.

Pour en savoir plus sur la protection et l'optimisation des performances des applications web utilisant Amazon CloudFront, consultez la page Démarrer avec CloudFront.

Résolution des noms de domaine à la limite (BP3)

Amazon Route 53 est un service de système de noms de domaine (DNS) 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, le routage basé sur la latence, le Geo DNS et les contrôles et la surveillance de la santé qui vous permettent de contrôler la façon dont le service répond aux demandes DNS, d'améliorer les performances de votre application web et d'éviter les pannes de site.

Amazon Route 53 utilise des techniques comme le "shuffle sharding" et le "anycast striping", qui peuvent aider les utilisateurs à accéder à votre application même si le service DNS est visé par une attaque DDoS. Avec le shuffle sharding, chaque serveur de nom de votre ensemble de délégation correspond à un ensemble unique d'emplacements de bordure et de chemins d'accès à Internet. Cela permet une plus grande tolérance aux pannes et minimise les chevauchements entre les clients. Si un serveur de noms de l'ensemble de délégation est indisponible, les utilisateurs peuvent réessayer et recevoir une réponse d'un autre serveur de noms situé à un autre emplacement. Le striping Anycast permet à chaque requête DNS d'être servie par l'emplacement le plus optimal, répartissant ainsi la charge du réseau et réduisant la latence DNS. Les utilisateurs obtiennent ainsi une réponse plus rapide. En outre, Amazon Route 53 peut détecter des anomalies dans la source et le volume des requêtes DNS, et

donner la priorité aux requêtes des utilisateurs qui sont connues pour être fiables.

Pour en savoir plus sur l'utilisation d'Amazon Route 53 pour diriger les utilisateurs vers votre application, voir Démarrer avec Amazon Route 53.

Application Layer Defense (BP1, BP2, BP6)

Bon nombre des techniques abordées dans ce document sont efficaces pour atténuer l'impact des attaques DDoS de la couche infrastructurelle sur la disponibilité de votre application. Pour se défendre également contre les attaques de la couche application, il faut mettre en place une architecture qui permet de détecter spécifiquement, d'évoluer pour absorber et bloquer les demandes malveillantes. Il s'agit là d'une considération importante, car les systèmes d'atténuation des DDoS basés sur le réseau sont généralement inefficaces pour atténuer les attaques complexes de la couche applicative.

Détecter et filtrer les requêtes web malveillantes (BP1, BP2)

Lorsque votre application fonctionne sur AWS, vous pouvez utiliser à la fois Amazon CloudFront et AWS WAF pour vous défendre contre les attaques DDoS de la couche application.

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

En utilisant AWS WAF, vous pouvez configurer des listes de contrôle d'accès au web (Web ACL) sur vos distributions CloudFront ou vos équilibreurs de charge d'application pour filtrer et bloquer les demandes en fonction des signatures de demande. Chaque ACL Web est constituée de règles que vous pouvez configurer pour faire correspondre une chaîne ou une regex à un ou plusieurs attributs de la requête, tels que l'URI, la chaîne de requête, la méthode HTTP ou la clé d'en-tête. De plus, en utilisant les règles basées sur le taux de l'AWS WAF, vous pouvez automatiquement bloquer 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 offensants recevront des réponses d'erreur "403 Interdit" et resteront bloquées jusqu'à ce que le taux de demandes descende en dessous du seuil. Cette mesure est utile pour atténuer les attaques de flood HTTP qui sont déguisées en trafic web régulier.

Pour bloquer les attaques provenant d'adresses IP malveillantes connues, vous pouvez créer des règles à l'aide des conditions de correspondance d'adresses IP ou utiliser les règles gérées pour les AWS WAF proposées par les vendeurs sur la place de marché AWS qui bloqueront les

adresses IP malveillantes spécifiques figurant dans les listes de réputation IP. AWS WAF et Amazon CloudFront vous permettent également de définir des restrictions géographiques pour bloquer ou mettre sur liste blanche les demandes provenant de certains pays. Cela peut aider à bloquer les attaques provenant de lieux géographiques où vous ne pensez pas pouvoir server les utilisateurs.

Pour vous aider à identifier les demandes malveillantes, vous pouvez consulter les journaux de votre serveur web ou utiliser la fonction "Sampled Requests" du WAF qui fournit des détails sur les demandes qui ont correspondu à l'une de vos règles AWS WAF pendant une période de temps au cours des 3 dernières heures. Vous pouvez utiliser ces informations pour identifier des 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 d'interrogation aléatoire, vous pouvez décider de désactiver l'interrogation de transmission de chaînes de caractères dans Amazon CloudFront. Cela peut être utile pour atténuer une attaque de cache contre votre origine.

Si vous êtes abonné à AWS Shield Advanced, vous pouvez faire appel à la DDoS Response Team (DRT) d'AWS pour vous aider à créer des règles visant à atténuer une attaque qui nuit à la disponibilité de votre application. La DRT ne peut obtenir qu'un accès limité à votre compte, et seulement avec votre autorisation explicite. Pour plus d'informations, voir la section Assistance plus loin dans ce document.

L'utilisation de AWS Firewall Manager peut aider à simplifier la gestion des règles AWS WAF dans votre organisation. En utilisant AWS Firewall Manager, vous pouvez activer AWS WAF sur de nombreux comptes et ressources, notamment en créant des règles qui sont automatiquement appliquées aux comptes existants ou nouveaux de votre organisation.

Pour en savoir plus sur l'utilisation de la géo-restriction pour limiter l'accès à votre distribution Amazon CloudFront, consultez la section Restriction de la distribution géographique de votre contenu.

Pour en savoir plus sur l'utilisation de l'AWS WAF, voir Démarrer avec l'AWS WAF et voir un échantillon des demandes Web que CloudFronta transmises à l'AWS WAF.

Pour en savoir plus sur la configuration des règles basées sur les tarifs, voir Protéger les sites et services web à l'aide de règles basées sur les tarifs pour les AWS WAF.

Pour savoir comment gérer le déploiement des règles WAF sur vos ressources AWS avec le gestionnaire de pare-feu AWS, consultez la section Démarrage avec le gestionnaire de pare-feu AWS.

Échelle d'absorption (BP6)

Une autre façon d'atténuer les attaques de la couche application est d'opérer à l'échelle. Si vous avez des applications web, vous pouvez utiliser ELB pour distribuer le trafic à un certain nombre d'instances d'Amazon EC2 qui sont surprovisionnées ou configurées pour s'étendre

automatiquement. Ces instances peuvent gérer les brusques hausses de trafic qui se produisent pour n'importe quelle raison, y compris une foule flash ou une attaque DDoS de la couche applicative. Vous pouvez configurer les alarmes d'Amazon CloudWatch pour qu'elles déclenchent une mise à l'échelle automatique afin d'augmenter automatiquement la taille de votre flotte Amazon EC2 en réponse aux événements que vous définissez. Cela permet de protéger la disponibilité des applications en cas d'augmentation inattendue du volume des demandes. Si vous utilisez Amazon CloudFront ou l'Application Load Balancer (ALB) avec votre application, la négociation TLS est gérée par la distribution ou l'équilibreur de charge. Cela permet de protéger vos instances contre les attaques basées sur le TLS en les adaptant pour traiter les demandes légitimes ainsi que les attaques d'abus du TLS.

Pour en savoir plus sur l'utilisation d'Amazon CloudWatch pour invoquer la mise à l'échelle automatique, voir Surveiller vos instances etgroupes de mise à l'échelle automatique en utilisant Amazon CloudWatch.

Réduction des surfaces d'attaque

Une autre considération importante lors de l'architecture d'une solution AWS est de limiter les possibilités qu'a un attaquant de cibler votre application. Par exemple, si vous ne vous attendez pas à ce que les utilisateurs interagissent directement avec certaines ressources, vous pouvez vous assurer que ces ressources ne sont pas accessibles depuis l'internet. De même, si vous ne vous attendez pas à ce que les utilisateurs ou les applications externes communiquent avec votre application sur certains ports ou protocoles, vous pouvez faire en sorte que ce trafic ne soit pas accepté.

Ce concept est connu sous le nom de *réduction de la surface d'attaque*. Dans cette section, nous fournissons les meilleures pratiques pour vous aider à réduire votre surface d'attaque et à limiter l'exposition de votre application à Internet. Les ressources qui ne sont pas exposées à l'internet sont plus difficiles à attaquer, ce qui limite les options dont dispose un attaquant pour cibler la disponibilité de votre application.

Obscurcir les ressources de l'AWS (BP1, BP4, BP5)

En général, les utilisateurs peuvent utiliser rapidement et facilement une application sans avoir besoin que les ressources de l'AWS soient entièrement exposées à l'internet. Par exemple, lorsque vous avez des instances d'Amazon EC2 derrière un ELB, il n'est pas nécessaire que les instances elles-mêmes soient accessibles au public. Vous pourriez plutôt fournir aux utilisateurs un accès à l'ELB sur certains ports TCP et permettre uniquement à l'ELB de communiquer avec les instances. Pour ce faire, vous pouvez configurer des groupes de sécurité et des listes de contrôle d'accès au réseau (NACL) dans votre Amazon Virtual Private Cloud (VPC). Amazon VPC vous permet de fournir une section logiquement isolée du nuage AWS où vous pouvez lancer des ressources AWS dans un réseau virtuel que vous définissez.

Les groupes de sécurité et les NACL sont similaires en ce sens qu'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 NACL offrent des capacités similaires au niveau du sous-réseau de la VPC. L'utilisation des groupes de sécurité ou des NACL n'entraîne pas de frais supplémentaires.

Groupes de sécurité et listes de contrôle d'accès aux réseaux (NACL) (BP5)

Vous pouvez spécifier des groupes de sécurité lorsque vous lancez une instance, ou vous pouvez associer l'instance à un groupe de sécurité ultérieurement. Tout le trafic internet vers

un groupe de sécurité est implicitement refusé, sauf si vous créez une règle d'autorisation pour permettre le trafic. Par exemple, si vous avez une application web qui utilise un ELB et un certain nombre d'instances Amazon EC2, vous pouvez décider de
créer un groupe de sécurité pour l'ELB ("ELB security group") et un autre pour les instances ("web application server security group"). Vous pouvez ensuite créer une règle d'autorisation pour permettre le trafic internet

au groupe de sécurité ELB, et une autre règle pour autoriser le trafic du groupe de sécurité ELB vers le groupe de sécurité du serveur d'application 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 de connaître et d'influencer votre application.

Lorsque vous créez des NACL, vous pouvez spécifier des règles d'*autorisation* et de *refus.* Cela est utile si vous souhaitez refuser explicitement certains types de trafic à votre application. Par exemple, vous pouvez définir des adresses IP (sous forme de plages CIDR), des protocoles et des ports de destination auxquels l'accès à l'ensemble du sous-réseau est refusé. Si votre application n'est utilisée que pour le trafic TCP, vous pouvez créer une règle pour *refuser* tout le trafic UDP, ou vice versa. Cette option est utile pour répondre aux attaques DDoS car elle vous permet de créer vos propres règles pour atténuer l'attaque lorsque vous connaissez les IP sources ou une autre signature.

Si vous êtes abonné à AWS Shield Advanced, vous pouvez enregistrer les IP élastiques (EIP) en tant que ressources protégées. Les attaques DDoS contre les EIP qui ont été enregistrés comme ressources protégées sont détectées plus rapidement, ce qui peut permettre de les atténuer plus rapidement. Lorsqu'une attaque est détectée, les systèmes d'atténuation des DDoS lisent le NACL qui correspond à l'EIP ciblé et l'appliquent à la frontière du réseau AWS. Cela réduit considérablement le risque d'impact d'un certain nombre d'attaques DDoS de la couche infrastructurelle.

Pour en savoir plus sur la configuration des groupes de sécurité et des NACL afin d'optimiser la résilience aux attaques DDoS, voir Comment aider à se préparer aux attaques DDoSen réduisant votre surface d'attaque.

Pour en savoir plus sur l'utilisation de AWS Shield Advanced avec les EIP comme ressources protégées, voir Activer et configurer AWS Shield Advanced.

Protéger vos origines (BP1, BP5)

Si vous utilisez Amazon CloudFront avec une origine qui se trouve à l'intérieur de votre VPC, vous devez utiliser une fonction Lambda de AWS pour mettre automatiquement à jour les règles de votre groupe de sécurité afin de n'*autoriser* que le trafic Amazon CloudFront. Cela améliore la sécurité de votre origine en garantissant que les utilisateurs malveillants ne peuvent pas contourner Amazon CloudFront et AWS WAF lorsqu'ils accèdent à votre application web.

Pour en savoir plus sur la façon de protéger votre origine en mettant automatiquement à jour vos groupes de sécurité, voir Comment mettre automatiquement à jour vos groupes de sécurité pour Amazon CloudFront et AWS WAF en utilisant AWS Lambda.

Vous pouvez également vous assurer que seule votre distribution Amazon CloudFront peut transmettre les demandes à votre origine. Avec les en-têtes de demande de bout en bout, vous pouvez ajouter ou remplacer la valeur des en-têtes de demande existants lorsqu'Amazon CloudFront transmet des demandes à votre origine. Vous pouvez utiliser l'en-tête *X-Shared-Secret* pour vous aider à valider que les demandes faites à votre origine ont été envoyées depuis Amazon CloudFront.

Pour en savoir plus sur la protection de votre origine avec un en-tête *X-Shared-Secret*, consultez la rubrique Transfert des en-têtes personnalisés vers votre origine.

Protection des points terminaux des API (BP4)

En général, lorsque vous devez exposer une API au public, il y a un risque que le frontend de l'API soit visé par une attaque DDoS. Pour réduire ce risque, vous pouvez utiliser la passerelle Amazon API comme "porte d'entrée" vers des applications fonctionnant sur Amazon EC2, AWS Lambda ou ailleurs. En utilisant Amazon API Gateway, vous n'avez pas besoin de vos propres serveurs pour le frontend API et vous pouvez obscurcir d'autres composants de votre application. En rendant plus difficile la détection de votre
vous pouvez contribuer à empêcher que ces ressources AWS ne soient la cible d'une attaque DDoS.

Lorsque vous utilisez la passerelle Amazon API, vous pouvez choisir entre deux types de points d'extrémité API. Le premier est l'option par défaut : il s'agit de terminaux API optimisés en périphérie, accessibles par le biais d'une distribution Amazon CloudFront. La distribution est toutefois créée et gérée par la passerelle API, de sorte que vous n'avez pas de contrôle sur elle. La deuxième option consiste à utiliser un point d'extrémité d'API régional auquel on accède depuis la même région AWS que celle dans laquelle votre API REST est déployée. Nous vous recommandons d'utiliser le second type de terminal, puis de l'associer à votre propre distribution Amazon CloudFront. Ce faisant, vous avez le contrôle de la distribution Amazon CloudFront et la possibilité d'utiliser AWS WAF pour la protection de la couche applicative.

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

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

  Pour en savoir plus sur la création d'API avec Amazon API Gateway, voir Démarrer avec Amazon API Gateway.

Techniques opérationnelles

Les techniques d'atténuation présentées dans ce document vous aident à concevoir des applications qui sont intrinsèquement résistantes aux attaques DDoS. Dans de nombreux cas, il est également utile de savoir quand une attaque DDoS vise votre application afin de pouvoir prendre des mesures d'atténuation. Si vous êtes victime d'une attaque, vous pouvez également bénéficier d'une aide pour évaluer la menace et revoir l'architecture de votre application, ou vous pouvez demander une autre assistance. Cette section traite des meilleures pratiques pour gagner en visibilité en cas de comportement anormal, d'alerte et d'automatisation, et pour faire appel à AWS pour une assistance supplémentaire.

Visibilité

Lorsqu'une mesure opérationnelle clé s'écarte considérablement de la valeur attendue, un attaquant peut tenter de cibler la disponibilité de votre application. Si vous connaissez bien le comportement normal de votre application, vous pouvez agir plus rapidement lorsque vous détectez une anomalie.

En utilisant Amazon CloudWatch, vous pouvez surveiller les applications que vous exécutez sur AWS. Par exemple, vous pouvez collecter et suivre des mesures, collecter et surveiller des fichiers journaux, régler des alarmes et répondre automatiquement aux changements de vos ressources AWS.

Si vous avez conçu votre application en suivant l'architecture de référence résistante aux DDoS, les attaques de la couche infrastructurelle commune seront bloquées avant d'atteindre votre application. Si vous êtes abonné à AWS Shield Advanced, vous avez accès à un certain nombre de mesures CloudWatch qui peuvent indiquer que votre application est visée. Par exemple, vous pouvez configurer des alarmes pour vous avertir lorsqu'une attaque DDoS est en cours, afin que vous puissiez vérifier votre
et décider d'engager ou non la DRT. Vous pouvez configurer la métrique DDoSDetected pour vous indiquer si une attaque a été détectée. Si vous souhaitez être alerté en fonction du volume de l'attaque, vous pouvez également utiliser les mesures DDoSAattackBitsPerSecond, DDoSAattackPacketsPerSecond ou DDoSAattackRequestsPerSecond. Vous pouvez surveiller ces mesures en intégrant Amazon CloudWatch à vos propres outils ou en utilisant des outils fournis par des tiers, tels que Slack ou PagerDuty.

Une attaque de la couche application peut faire monter de nombreuses mesures d'Amazon CloudWatch. Si vous utilisez AWS WAF, vous pouvez utiliser CloudWatch pour surveiller et alerter sur l'augmentation des demandes que vous avez définies dans

WAF à autoriser, à compter ou à bloquer. Cela vous permet de recevoir une notification si le niveau de trafic dépasse ce que votre application peut traiter. Vous pouvez également utiliser les mesures Amazon CloudFront, Amazon Route 53, ALB, NLB, Amazon EC2 et Auto Scaling qui sont suivies dans CloudWatch pour détecter les changements qui peuvent indiquer une attaque DDoS

Pour une description des mesures d'Amazon CloudWatch qui sont couramment utilisées pour détecter et réagir aux attaques DDoS, voir le tableau suivant (Tableau 3).

Sujet	Métrique	Description
AWS Shield Advanced	DDoS détecté	Indique un événement DDoS pour un nom de ressource amazonienne (ARN) spécifique.
AWS Shield Advanced	Attaque DDoS : nombre de bits par seconde	Le nombre d'octets observés lors d'un événement DDoS pour un nom de ressource amazonienne (ARN) spécifique. Cette mesure n'est disponible que pour les événements DDoS de couche 3/4.
AWS Shield Advanced	Paquets d'attaque DDoS par seconde	Le nombre de paquets observés lors d'un événement DDoS pour un nom de ressource amazonienne (ARN) spécifique. Cette mesure n'est disponible que pour les événements DDoS de couche 3/4.
AWS Shield Advanced	Demandes d'attaques DDoS par seconde	Le nombre de demandes observées lors d'un événement DDoS pour un nom de ressource amazonienne (ARN) spécifique. Cette mesure n'est disponible que pour les événements DDoS de couche 7 et n'est rapportée que pour les événements de couche 7 les plus significatifs.
AWS WAF	Demandes acceptées	Le nombre de requêtes web autorisées.
AWS WAF	Demandes bloquées	Le nombre de requêtes web bloquées.
AWS WAF	Demandes comptabilisées	Le nombre de demandes web comptabilisées.
Amazon CloudFront	Demandes	Le nombre de demandes HTTP/S
Amazon CloudFront	Taux d'erreur total	Le pourcentage de toutes les demandes pour lesquelles le code d'état HTTP est 4xx ou 5xx.
Route de l'Amazone 53	État d'avancement du bilan de santé	Le statut du point final du bilan de santé.
ALB	Nombre de connexions actives	Le nombre total de connexions TCP simultanées qui sont actives des clients vers l'équilibreur de charge, et de l'équilibreur de charge vers les cibles.
ALB	Consommation d'ULC	Le nombre d'unités de capacité de l'équilibreur de charge (LCU) utilisé par votre équilibreur de charge.

Sujet	Métrique	Description
ALB	HTTPCode_ELB_4XX_Compte HTTPCode_ELB_5XX_Compte	Le nombre de codes d'erreur client HTTP 4xx ou 5xx générés par l'équilibreur de charge.
ALB	Le nombre de nouvelles connexions	Nombre total de nouvelles connexions TCP établies entre les clients et l'équilibreur de charge, et entre l'équilibreur de charge et les cibles.
ALB	Octets traités	Le nombre total d'octets traités par l'équilibreur de charge.
ALB	Nombre de connexions rejetées	Le nombre de connexions qui ont été rejetées parce que l'équilibreur de charge avait atteint son nombre maximum de connexions.
ALB	Demande de comptage	Le nombre de demandes qui ont été traitées.
ALB	Nombre d'erreurs de connexion à la cible	Le nombre de connexions qui n'ont pas été établies avec succès entre l'équilibreur de charge et la cible.
ALB	Temps de réponse cible	Le temps écoulé, en secondes, après que la demande ait quitté l'équilibreur de charge jusqu'à ce qu'une réponse de la cible soit reçue.
ALB	Compter les hôtes malsains	Le nombre de cibles considérées comme malsaines.
NLB	Comptage des flux actifs	Le nombre total de flux TCP simultanés (ou connexions) des clients vers les cibles.
NLB	Consommation d'ULC	Le nombre d'unités de capacité de l'équilibreur de charge (LCU) utilisé par votre équilibreur de charge.
NLB	Nouveau décompte des flux	Le nombre total de nouveaux flux TCP (ou connexions) établis entre les clients et les cibles au cours de la période.
NLB	Octets traités	Le nombre total d'octets traités par l'équilibreur de charge, y compris les en-têtes TCP/IP.
Mise à l'échelle automatique	Taille maximale du groupe	La taille maximale du groupe de mise à l'échelle automatique
Amazone EC2	Utilisation du CPU	Le pourcentage d'unités de calcul EC2 allouées qui sont actuellement utilisées.
Amazone EC2	Réseau en	Le nombre d'octets reçus par l'instance sur toutes les interfaces réseau.

Tableau 3 : Paramètres recommandés pour l'Amazon CloudWatch.

Pour en savoir plus sur l'utilisation d'Amazon CloudWatch pour détecter les attaques DDoS sur votre application, consultez la section Démarrer avec Amazon CloudWatch.

L'AWS comprend plusieurs mesures et alarmes supplémentaires pour vous avertir d'une attaque et vous aider à surveiller les ressources de votre application. La console ou l'API AWS Shield fournit un résumé et des détails sur les attaques qui ont été détectées. En outre, le tableau de bord de l'environnement de la menace globale fournit des informations récapitulatives sur toutes les attaques DDoS qui ont été détectées par AWS. Cela peut être utile pour mieux comprendre les menaces DDoS sur une plus grande population d'applications, comprendre les tendances des attaques et les comparer avec les attaques que vous avez pu observer.

Les journaux de flux VPC sont 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 les journaux de flux réseau pour résoudre les problèmes de connectivité et de sécurité, et pour vous assurer que les règles d'accès au réseau fonctionnent comme prévu. En utilisant les journaux de flux VPC, vous pouvez capturer des informations sur le trafic IP qui va vers et depuis les interfaces réseau dans votre VPC.

Chaque enregistrement du journal des flux comprend 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 vous aider à identifier les anomalies du trafic réseau et pour identifier un vecteur d'attaque spécifique. Par exemple, la plupart des attaques par réflexion UDP ont des ports sources spécifiques, comme le port source 53 pour la réflexion DNS. Il s'agit d'une signature claire que vous pouvez identifier dans l'enregistrement du journal des flux. En réponse, vous pouvez choisir de bloquer le port source spécifique au niveau de l'instance ou de créer une règle NACL pour bloquer l'ensemble du protocole si votre application
ne l'exige pas.

Pour en savoir plus sur l'utilisation des VPC Flow Logs pour identifier les anomalies du réseau et les vecteurs d'attaque DDoS, voir VPC Flow Logs et VPC Flow Logs - Log et View Network Traffic Flows.

Soutien

Il est important de créer un plan de réponse aux attaques DDoS avant un événement réel. Les meilleures 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. Examinez les options de cette section afin de 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 vous aider à répondre aux questions ou aux défis spécifiques que vous vous posez.

Si vous exécutez des charges de travail de production sur AWS, envisagez de vous abonner au service d'assistance aux entreprises qui vous donne accès 24 heures sur 24 et 7 jours sur 7 à des ingénieurs d'assistance dans le nuage qui peuvent vous aider à résoudre les problèmes d'attaques DDoS. Si vous avez des charges de travail critiques, envisagez de souscrire au service d'assistance aux entreprises qui vous permet d'ouvrir des dossiers "critiques" et de recevoir la réponse la plus rapide d'un spécialiste de l'assistance aux entreprises.
Ingénieur principal de soutien au nuage.

Si vous êtes abonné à AWS Shield Advanced et que vous êtes également abonné à l'assistance aux entreprises ou à l'assistance aux entreprises, vous pouvez passer à l'équipe d'intervention DDoS (DRT) d'AWS si vous avez un événement lié à un DDoS qui a une incidence sur la disponibilité de votre application. Si la réactivité de votre application est dégradée à cause d'une attaque DDoS, vous pouvez contacter en direct l'équipe d'assistance AWS. Une autre option consiste à utiliser la fonction Lambda d'engagement du bouclier AWS pour établir plus rapidement le contact avec la DRT. Par exemple, vous pouvez utiliser un bouton IoT AWS pour déclencher la fonction Lambda AWS si vous êtes dans une situation d'urgence. Lorsque vous appuyez sur le bouton, un dossier est automatiquement ouvert avec l'assistance AWS et la DRT est immédiatement avertie. Vous recevez une réponse directe pour votre cas qui comprend un pont de conférence Amazon Chime que vous pouvez rejoindre pour interagir avec AWS Support et DRT. Le Lambda d'engagement du bouclier AWS peut être utilisé avec tout déclencheur supporté par le Lambda AWS.

Pour en savoir plus sur l'engagement rapide de la DRT en utilisant une fonction Lambda de l'AWS, voir Engagement de bouclier Lambda de l'AWS.

DRT n'a généralement pas accès à votre compte AWS ou à vos demandes échantillonnées AWS WAF. Vous pouvez autoriser DRT à accéder à AWS WAF, AWS Shield et aux opérations API connexes sur votre compte à partir de la console AWS Shield ou de l'API. Par exemple, vous pouvez autoriser DRT à visualiser vos demandes échantillonnées ou à placer des règles pour aider à atténuer une attaque DDoS de la couche application. Vous pouvez également autoriser DRT à accéder aux seaux Amazon S3 que vous spécifiez. Par exemple, vous pouvez avoir un seau dans lequel vous stockez les journaux de requêtes web et vous souhaitez que DRT y ait accès pour les analyser pendant une attaque. DRT n'accédera à votre compte ou n'y apportera

des modifications que pendant une escalade et tout changement sera soumis à votre consentement. Pour en savoir plus sur l'octroi d'un accès limité à DRT, voir Autoriser l'équipe d'intervention DDoS.

Dans certains cas, la DRT peut se renseigner sur une attaque DDoS et vous engager de manière proactive. S'il existe des points de contact spécifiques qui devraient être engagés lors d'une escalade déclenchée par DRT, vous pouvez les ajouter dans la console AWS Shield en cliquant sur "Summary", puis sur "Edit" sous l'onglet
Section "Contacts supplémentaires".

Conclusion

Les meilleures pratiques décrites dans ce document peuvent vous aider à mettre en place une architecture résistante aux attaques DDoS qui peut protéger la disponibilité de vos applications en empêchant de nombreuses attaques DDoS courantes au niveau de l'infrastructure et de la couche applicative. La mesure dans laquelle vous suivez ces meilleures pratiques lorsque vous concevez votre application influencera le type, le vecteur et le volume des attaques DDoS que vous pouvez atténuer. Vous pouvez intégrer la résilience sans souscrire à un service d'atténuation des attaques DDoS. Vous pouvez également choisir de vous abonner à AWS Shield Advanced pour bénéficier de fonctions supplémentaires de support, de visibilité, d'atténuation et de protection des coûts qui protègent davantage une architecture d'application déjà résiliente.

Pour en savoir plus sur l'atténuation des DDoS et les meilleures pratiques en matière de résilience aux DDoS sur les AWS, voir l'annexe A : Ressources supplémentaires.

Contributeurs

Les personnes et organisations suivantes ont contribué à ce document :

- Andrew Kiggins, architecte des solutions AWS
- Jeffrey Lyon, Protection du périmètre AWS
- Achraf Souk, architecte des solutions AWS
- Tino Tran, architecte des solutions AWS
- Yoshihisa Nakatani, architecte des solutions AWS

Révision des documents

Date	Description
Juin 2018	Mise à jour pour inclure AWS Shield, les fonctionnalités AWS WAF, AWS Firewall Manager et les meilleures pratiques connexes.
Juin 2016	Ajout d'un guide d'architecture prescriptif et mise à jour pour inclure le WAF AWS.

Date	Description
Juin 2015	Première publication.

Annexe A : Ressources supplémentaires

Vous pouvez utiliser les ressources suivantes pour en savoir plus sur l'atténuation des DDoS et les meilleures pratiques de résilience aux DDoS avec l'AWS :

- Meilleures pratiques pour l'atténuation des DDoS sur les AWS
- SID216 - re:Invent 2017 : Protection des applications natives des nuages : Sécurité des applications web à Pearson
- SID324 - re:Invent 2017 : Automatisation de la réponse DDoS dans le nuage
- CTD304 - re:Invent 2017 : Le voyage de gestion du Dow Jones et du Wall Street Journal Des pics de trafic tout en atténuant les menaces de DDoS et de la couche applicative
- CTD310 - re:Invent 2017 : Vivre sur le fil du rasoir, c'est plus sûr que vous ne le pensez ! Bâtiment Fort avec Amazon CloudFront, AWS Shield et AWS WAF