实现 DDoS 弹性的 AWS 最佳实践

2015 年 6 月

版权声明

本文档仅供参考。本文档只体现自其发行之日起 AWS 当前的产品服务和实践，如有变更，恕不另行通知。客户负责对本文档的信息以及对 AWS 产品或服务的任何使用情况进行自我独立的评估，每项产品或服务均按“原样”提供，没有任何形式的保证，无论是明示的还是暗示的。本文档不形成 AWS、其附属公司、供应商或许可方的任何保证、表示、合同承诺、条件或担保。AWS 对其客户承担的责任和义务受 AWS 协议的制约，本文档不是 AWS 与客户直接协议的一部分，也不构成对该协议的修改。

目录

摘要

本文是面向想要改进其在 Amazon Web Services (AWS) 上运行的应用程序的弹性，以抵御分布式拒绝服务攻击的客户。本文概述了分布式拒绝服务攻击、有助于保持可用性的技术以及参考架构，以提供旨在改进弹性的架构指导。

本文面向 IT 决策者和安全人员，并假定这些人员熟悉网络、安全和 AWS 方面的基本概念。每个章节都包含了指向 AWS 文档的链接，这些文档提供有关如何执行所列任务的详细信息。此外，您还可以观看 AWS re:Invent 会议视频 Sec305 和 SEC307 以了解更多信息。

简介

拒绝服务 (DoS) 攻击是指尝试使您的网站或应用程序无法为您的最终用户提供服务的攻击形式。为达到这一目的，攻击者会运用多种耗用网络或其他资源的技术手段来中断最终用户的合法访问。最简单的 DoS 攻击形式是攻击者本人通过单一主机对目标实施攻击，如图 1 所示。

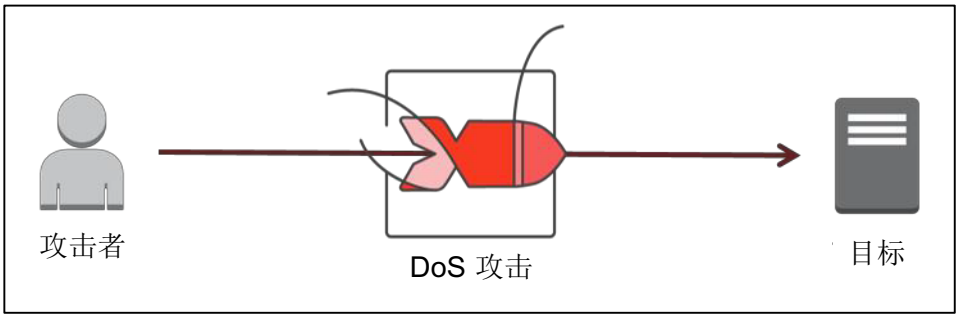

图 1：DoS 攻击示意图

分布式拒绝服务攻击

在分布式拒绝服务 (DDoS) 攻击形式中，攻击者将借助多台主机（可能遭到一组协作者的盗用或控制）来策划对目标展开攻击。如图 2 所示，在 DDoS 攻击中，每个协作者或遭到盗用的主机均参与攻击活动，从而生成海量的数据包或请求来“淹没”预定目标。

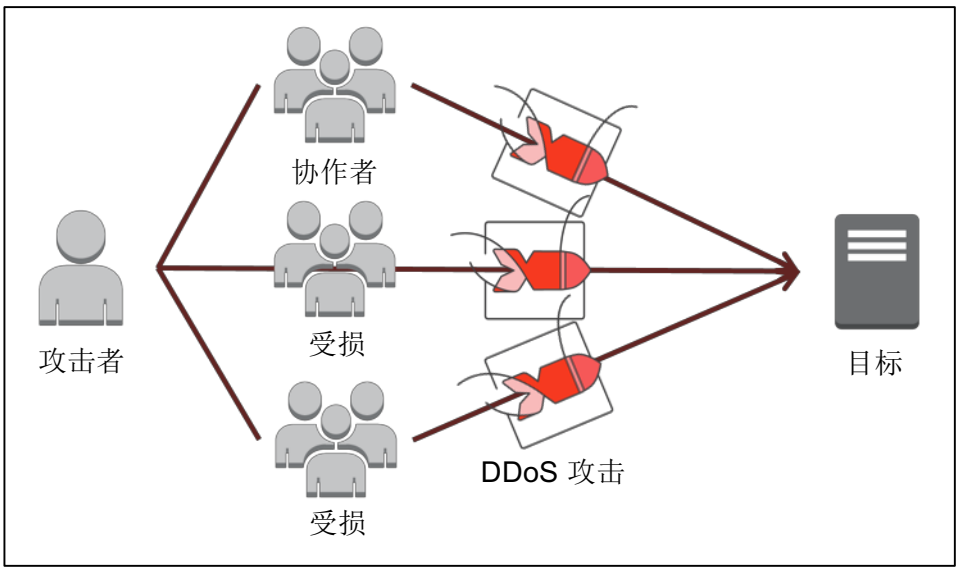

图 2：DDoS 攻击示意图

攻击者“淹没”目标的方法有很多种。比如，攻击者可以通过组合使用反射和放大技术或利用大型僵尸网络来产生海量的数据包。反射攻击涉及诱发从服务器到假冒 IP 地址的响应（其中，被入侵的服务器起着反射器的作用）。

放大攻击的原理则是攻击者通过发送较小的数据包或请求来诱发大量的响应。放大系数（请求大小与响应大小的比值）随所用的协议（如 DNS、NTP、SSDP）而变。例如，DNS 的平均放大系数介于 28 到 54 之间 — 也就是说，攻击者向 DNS 服务器发送 64 字节的请求负载可以生成 3,456 字节的非必要流量。[1]

将反射攻击与放大攻击结合起来可使被盗用的服务器发送与原始请求不成比例的响应。如图 3 所示，通过结合使用这些方法，攻击者可将来自有限数量主机的少量带宽转变成极大的流量来“轰击”目标。

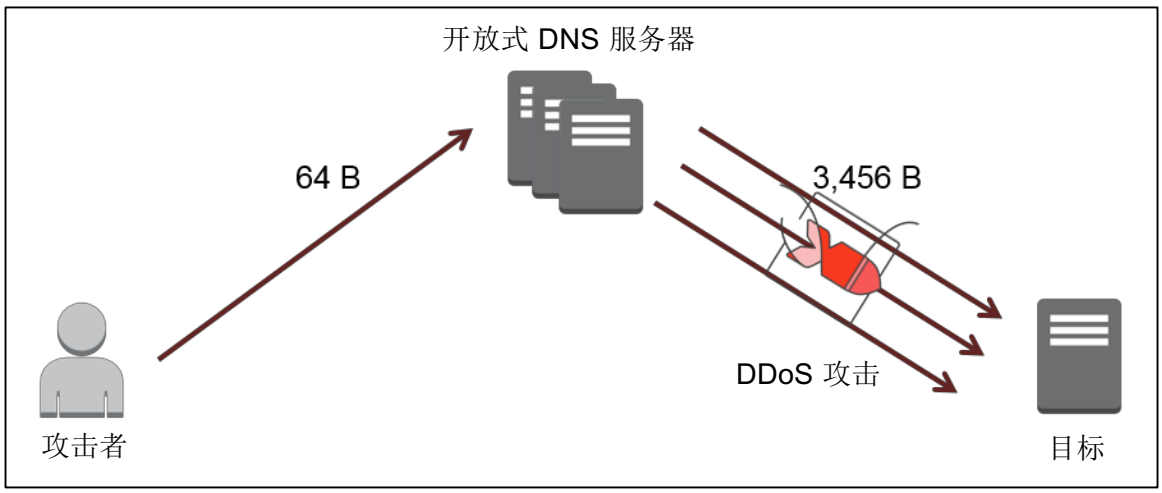

图 3：DDoS 泛洪、反射和放大攻击

当不必要的流量到达目标时，流量将穿过多层网络硬件、操作系统和应用程序层，并可能耗尽这些层中任意一层的资源。根据所消耗的资源，这种攻击可划分为带宽耗尽（如 UDP 泛洪）、协议耗尽（如 SYN 泛洪）或应用程序耗尽（如 HTTP GET/POST 泛洪）。

[1] https://www.us-cert.gov/ncas/alerts/TA14-017A

不管攻击类型如何，DDoS 攻击的核心都是威胁可用性，因为攻击者的目标是使合法的最终用户无法使用资源。因此，您可以利用 AWS 中的故障转移功能来降低您遭遇 DDoS 攻击时导致的可用性问题的风险。

根据所用的服务和服务的配置方式，抵御 DDoS 攻击的弹性程度可能会有所差异；因此，此处描述的技术不保证实现特定的可用性水平。使用本文中罗列的技术时，请记住，AWS 是基于使用情况而定价的。

缓解技术

通常，DDoS 安全涉及创建过滤器和屏障，来阻止攻击者达到其目的。除了过滤和拦截技术，您还可以采用能够随环境变化而扩展的灵活架构。本文其余部分讨论的是可用于降低遭遇 DDoS 攻击风险的五大技术：

- 最大限度地缩小攻击面
- 做好扩展准备，以吸收攻击
- 保护公开的资源
- 了解正常行为
- 制订攻击防范计划

本文讨论各项技术并提供参考架构（如图 4 所示），以演示如何利用 AWS 构建弹性。您也可以遵循 AWS 架构中心中的一般准则来构建能够快速进行故障转移的系统（如容错能力和高可用性中所示）。

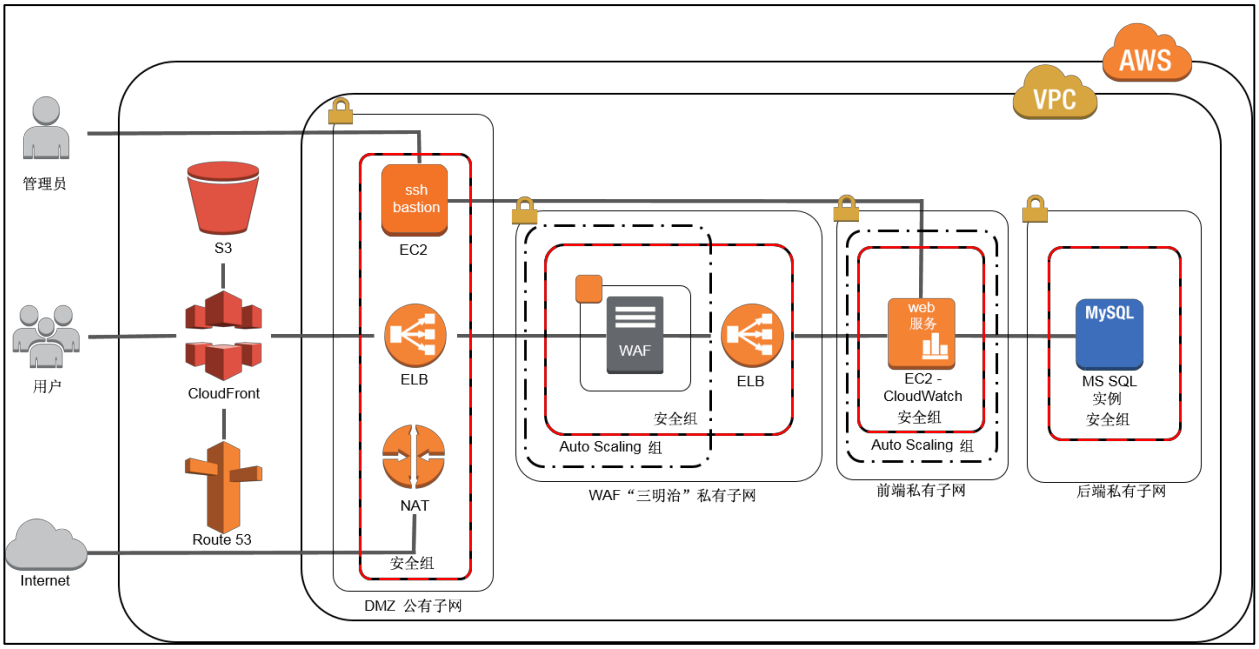

图 4：DDoS 弹性参考架构

最大限度地缩小攻击面

攻击面由允许访问您的应用程序的不同的 Internet 入口点组成。最大限度缩小攻击面的策略是：(a) 减少必要的 Internet 入口点的数量；(b) 消除非关键的 Internet 入口点；(c) 将最终用户流量与管理流量分离；(d) 混淆必要的 Internet 入口点，使不可信的最终用户无法访问它们；以及 (e) 隔断 Internet 入口点，以尽量减少攻击的影响程度。可通过 Amazon Virtual Private Cloud 来实现这一策略。

Amazon Virtual Private Cloud

Amazon Virtual Private Cloud (VPC) 让您可以在 AWS 内属于您自己的逻辑隔离区域中定义虚拟网络，我们称之为虚拟私有云。VPC 让您能够配置路由表、网络网关和安全设置，以便在安全的环境中启动 EC2 实例等 AWS 资源。

更重要的是，VPC 让您能够隐藏实例，使其对 Internet 不可见，以确保非公开实例仅在私有子网上可用，并且私有 DNS 条目只能由内部应用程序访问。这可通过创建安全组和网络访问控制列表 (ACL) 来实现。安全组充当关联的 EC2 实例的防火墙，网络 ACL 则充当关联子网的防火墙。您可以将安全组用作第一防御层来确保您的 VPC 实例的安全，并添加网络 ACL 作为第二防御层。有关安全组和网络 ACL 的更多信息，请参阅您的 VPC 中的安全性。

利用以下链接在 Amazon 管理控制台中配置 VPC。如果您不确定应在步骤 3 中选择何种实例，请先阅读名为 Amazon Elastic Compute Cloud 的章节，然后再执行后续步骤。

步骤 1：设置 VPC 和 Internet 网关

步骤 2：为您的 VPC 设置安全组

步骤 3：将实例启动到 VPC 中

步骤 4：为您的实例分配弹性 IP 地址

步骤 5：设置网络访问控制列表

您还可以遵循下列 Amazon VPC 常见配置情景的步骤执行操作。

情景 1：仅带有公有子网的 VPC

情景 2：带有公有子网和私有子网的 VPC

情景 3：带有公有子网和私有子网以及硬件 VPN 访问的 VPC

情景 4：仅带有私有子网和硬件 VPN 访问的 VPC

遵循这些链接中的说明执行操作后，您即可连接位于您 VPC 中的 EC2 实例。有关如何连接 Linux 实例的详细信息，请参阅*适用于 Linux 实例的 Amazon EC2 用户指南* 中的连接 Linux 实例。有关如何连接 Windows 实例的信息，请参阅*适用于 Microsoft Windows 实例的 Amazon EC2 用户指南* 中的连接 Windows 实例。

做好扩展准备，以吸收攻击

DDoS 攻击的重点在于规模。大多数攻击者通过发送应用程序无法容纳的流量水平来达到其目的。通过实施能够承受大规模攻击的架构，您可创造一种需要攻击者耗费更多时间和资源的屏障，从而使您的应用程序更具弹性。

在 AWS 中，您可以采取两种形式的扩展：水平扩展和垂直扩展。水平扩展通过向您的基础设施添加更多的实例或服务来实现。垂直扩展通过选择具有更多内存、CPU 和容量的实例来实现。利用这两种维度的扩展可提供抵御 DDoS 攻击的四大直接好处：

- 使攻击分散到更广阔的区域，最大限度地减小“轰击”半径。
- 攻击者必须消耗更多的资源才能扩大攻击规模。
- 扩展为您赢得了分析 DDoS 攻击并通过对抗措施予以回应的宝贵时间。
- 扩展针对其他故障情景提供了额外的冗余层。

在 DDoS 攻击方面，AWS 提供三种可用于扩展的方式：(1) 为您的应用程序选择合适的实例类型；(2) 配置 Elastic Load Balancing、Auto Scaling 等服务以实现自动扩展；以及 (3) 利用 Amazon CloudFront、Amazon Route 53 等 AWS 全球性服务固有的扩展性。

Amazon Elastic Compute Cloud

实例类型

Amazon Elastic Compute Cloud (EC2) 在云中提供可扩展的计算容量，而无需在硬件上提前投入。您可以借助 EC2 来启动被称作为实例的虚拟服务器，实例可以扩展或收缩，以应对流量峰值的变化。启动实例时，您指定的实例类型决定了用于您的应用程序的主机硬件组件（如计算机、内存、存储等）。

为提高应用程序的弹性，您应该选择能够扩展的实例类型，以支持您的应用程序及意料之外的流量峰值。某些针对成本而优化的 EC2 实例不提供保障性的网络资源，它们可能对 DDoS 攻击的可用性影响更为敏感。对于那些对基础设施至关重要的应用程序，您应选择 EBS 优化实例或具有 10 Gb 网络连接能力的实例类型作为主机。有关更多信息，请参阅 Amazon EC2 实例配置。

增强联网

对于 C3、C4、R3、D2 和 I2 实例，您可以启用增强联网功能，以提供更高的网络性能（每秒数据包数）。与传统实现方式相比，该功能使用了能够提供更高 I/O 性能和更低 CPU 利用率的网络虚拟化堆栈。借助增强联网功能，您的应用程序能够从帮助构建抵御 DDoS 攻击的弹性的功能中受益，如较高的每秒数据包数性能、低延迟联网和更高的可扩展性。

要启用增强联网功能，您必须启动 Amazon 系统映像 (AMI)，它是具有单根 I/O 虚拟化 (SR-IOV) 驱动程序的相应硬件辅助虚拟机 (HVM)。Amazon HVM Linux AMI 默认包含 SR-IOV 驱动程序。对于默认不包含 SR-IOV 驱动程序的 AMI，您需要下载并安装适当的驱动程序。有关如何针对以下 AMI 下载并启用增强联网功能的说明，请参考下面相应的链接。

在 Amazon Linux 上启用增强联网

在 Ubuntu 上启用增强联网

在其他 Linux 发行版上启用增强联网

在 Windows 上启用增强联网

Elastic Load Balancing

使用 Elastic Load Balancing (ELB) 管理您基础设施的流量时，您可获得将流量分布到跨多个可用区 (AZ) 的若干 EC2 实例的好处，从而最大限度地降低使用单一实例过载的风险。借助于 ELB，您可以根据需求的变化来添加和删除 EC2 实例，而不会中断整体流量。例如，当某个 EC2 实例发生故障时，ELB 会将流量自动重新路由至其余正在运行的 EC2 实例。当出现故障的 EC2 实例恢复后，ELB 将恢复前往该实例的流量。

此外，ELB 还提供单一管理点，并可作为防御网络攻击的第一道防线。您可以将自己所有的 EC2 实例都置于 ELB 之后，仅向 Internet 公开 ELB，这有助于最大限度地缩小攻击面。您可以将实例指向 ELB 并根据需要扩展（添加或删除容量），而无需管理每个单独的实例。搭配 VPC 使用时，您可以为 ELB 关联安全组和网络 ACL，以提供额外的弹性层。ELB 只支持有效的 TCP 请求，因此，UDP 和 SYN 泛洪之类的 DDoS 攻击无法到达您的实例。

下一章节将讲解如何配置两种基本的负载均衡器：面向 Internet 的负载均衡器和内部的负载均衡器。这两种负载均衡器（如图 4 所示）对于 Web 应用程序防火墙章节中详述的扩展第 7 层的保护来说是必不可少的。

步骤 1：在默认 VPC 中创建基本负载均衡器

步骤 2：在 Amazon VPC 中创建基本内部负载均衡器

对于您拥有 EC2 实例的其他区域重复这些步骤。

Auto Scaling

Auto Scaling 可帮助您维持应用程序的可用性，并允许您根据定义的条件自动扩展或收缩 EC2 容量。例如，您可以设置条件,以便在网络流量较高时（典型的 DDoS 攻击）向 Auto Scaling 组逐步添加新的实例。此外，您还可以设置条件在网络流量较低时以相同的增量削减实例。您可以使用 Amazon CloudWatch 触发扩展活动，并由 ELB 将流量分布到您 Auto Scaling 组中的实例。

在将第一个 Auto Scaling 组投入生产之前，您需要考虑一些行动方案。开始规划之前，请花些时间全面考察应用程序在 AWS 云中运行时的情况，并记录以下事项：

- 启动和配置您的服务器需要多长时间？*如果您的应用程序的启动时间超过五分钟，则我们建议让多个实例预先运行应用程序或降低触发扩展的阈值。*
- 哪些指标与应用程序的性能关系最密切？*DDoS 攻击的示例指标有* `CPUUtilization`、`NetworkIn` *和* `StatusCheckFailed`。
- 您可能要在 Auto Scaling 组中使用哪些现有资源（如 EC2 实例或 AMI）？*当应用程序遭受攻击时，您需要为 Auto Scaling 组使用相同类型的实例或更高的容量。*
- 您希望 Auto Scaling 组跨多少个可用区？*我们建议至少跨两个可用区。*
- 应以多快的速度扩展或收缩？*请记住，DDoS 攻击可能是一波波地来袭。估计您也不想在第一波攻击后就收缩，然后才发现不得不再次扩展。*
- Auto Scaling 组的最大 EC2 实例数是多少？*额外的实例可能会增加您的成本。在创建 Auto Scaling 策略时，您可以设置实例的最大数量。您还可以设置在达到该最大数量时报警。有关设置警报的步骤，请参阅 Amazon CloudWatch。*

您对基础设施和应用程序的了解越深，Auto Scaling 的实施效果就越好。当您掌握了有关基础设施和应用程序的足够信息后，就可以开始创建 Auto Scaling 组了。

步骤 1：创建启动配置

步骤 2：创建 Auto Scaling 组

步骤 3：验证 Auto Scaling 组

Amazon CloudFront

Amazon CloudFront 是一种内容分发网络 (CDN)，它与 Amazon 其他的 服务集成，提供一种将内容便捷、高效分发给最终用户的方式。CDN 充当您的源与最终用户之间的代理层。CDN 通过缓存内容和优化，从多个接入点连接 (PoP) 到您的站点来提高性能。它产生的效果是：不将所有流量请求发回到您的源，而是将请求分散到多个 PoP 上，并直接响应您的最终用户。

通过使用多个 PoP，Amazon CloudFront 具有的固有能力，可以将流量分散到多个位置，从而帮助缓解针对基础设施和一些应用程序层的 DDoS 攻击。在每个位置，AWS 都有多个用于提供容量和冗余的 Internet 连接，从而使得 Amazon CloudFront 能够隔离攻击流量，同时向合法的最终用户提供内容。

此外，Amazon CloudFront 还具有过滤功能，以确保只生成有效的 TCP 连接和 HTTP 请求，并丢弃无效的请求。这可减轻您的源在处理无效流量时（常用在 UDP 泛洪、SYN 泛洪、缓慢读取攻击中）的负担。

要使用 Amazon CloudFront，您可以创建分配并指定源（可以是 EC2 实例、S3 存储桶、ELB 或自定义 Web 服务器）。在您配置分配后，Amazon CloudFront 会开始响应最终用户请求，并缓存内容以提高性能。

以下说明介绍如何创建使用 ELB 作为源的 Amazon CloudFront 分配（如图 4 所示）。有关为不同的分配配置 Amazon CloudFront 的信息，请参阅使用 Amazon EC2 和其他自定义源的要求和建议。

步骤 1：创建 Web 分配

步骤 2：测试 Web 分配

Amazon Route 53

最常见的 DDoS 攻击目标之一是域名系统 (DNS)。DNS 通常被用来 定位和将域名转换为 IP 地址，因此，攻击者常将 DNS 视为单一故障点。例如，您的应用程序实际上可能正常运行，但如果您的应用程序的 DNS 发生故障，则最终用户就无法被正确地路由。这会造成应用程序不可用的效果。DNS 使用 UDP 协议响应查询，因此，该服务非常容易受到大规模的反射和放大攻击的影响。出于上述原因，您将需要资源来提高 DNS 的弹性。

Amazon Route 53 是一款高度可用且可扩展的 DNS 服务，旨在将最终用户路由到在 AWS 上内外运行的基础设施。Amazon Route 53 让您可以通过多种路由类型（包括基于延迟的路由、Geo DNS 和加权轮询算法）来管理全球流量。这些路由类型还可以与 Route 53 故障转移结合使用，以实现低延迟的容错架构。

Amazon Route 53 有两项相互协作的功能：随机分区和选播路由。甚至在遭受 DDoS 攻击的情况下，它们也能确保最终用户访问到您的应用程序。

随机分区

随机分区在概念上类似于数据库分区，后者是将数据的水平分区分散到单独的数据库服务器上，以分散负载并提供冗余。同样，Amazon Route 53 采用随机分区来将 DNS 请求分散到多个 PoP 上，从而为您的应用程序提供多个路径和路由。

选播路由

选播路由通过多个 PoP 广播相同的 IP 地址来提高冗余。一旦 DDoS 攻击“淹没”了某个终端节点，随机分区会隔离故障，同时提供到您的基础设施的其他路由。

此外，Amazon Route 53 还提供 DNS 故障转移功能的健康检查，以确保 DNS 查询只使用健康的资源。例如，假设 example.com 托管在 10 个实例上，其中，有两个实例位于全世界不同的可用区。您可以配置 Amazon Route 53 来检查这些实例的健康状况，并只使用健康的实例响应 example.com 的 DNS 查询。您可以使用 Amazon Route 53 别名、加权轮询算法、基于延迟的路由、Geo DNS 和故障转移资源记录集来设置各种各样的故障转移配置。

要将 Amazon Route 53 用于 DNS，您可以将域名和子域注册或迁移到 Amazon Route 53。本章节将使用 Amazon CloudFront 分配作为 Amazon Route 53 的 ALIAS 记录集。有关将流量路由到其他服务的步骤，请参阅将查询路由到 AWS 资源中所列的步骤。

步骤 1：注册域名并将 Amazon Route 53 配置为 DNS 服务

步骤 2：将查询路由到 Amazon CloudFront 分配（仅限公有托管区域）

步骤 3：创建健康检查和 DNS 故障转移

有关使用 Amazon Route 53 的更多信息，请参阅以下链接。

- 将现有域的 DNS 服务迁移到 Amazon Route 53
- 创建使用 Amazon Route 53 作为 DNS 服务的子域，而不迁移父域
- 将子域的 DNS 服务迁移到 Amazon Route 53，而不迁移父域

保护公开的资源

如果您无法消除到您应用程序的 Internet 入口点，则您需要采取其他措施以在不中断合法最终用户流量的前提下限制访问和保护这些入口点。可提供该控制和灵活性的三种资源是 Amazon CloudFront、Amazon Route 53 和 Web 应用程序防火墙 (WAF)。

Amazon CloudFront

Amazon CloudFront 提供了两种机制来限制对内容的访问：*地理限制* 和 *来源访问标识*。

地理限制

Amazon CloudFront 支持地理限制（也称作地理阻止），可阻止特定地理位置对您内容的访问。最终用户请求您的内容时，Amazon CloudFront 通常会提供请求的内容，而不考虑请求的来源。但是，如果您的最终用户群位于特定的国家/地区，则您可以使用地理限制来降低将您的服务暴露给其他国家或地区（通常不在您的最终用户的区域？）的程度。地理限制支持以下场景：

- 基于批准的国家或地区的白名单来允许对内容的访问
- 基于禁止的国家或地区的黑名单来阻止对内容的访问

要使用地理限制，您可以使用内置的 Amazon CloudFront 功能或使用更详细的高于国家或地区级别的第三方地理位置服务。要使用内置的 Amazon CloudFront 功能，请遵循限制内容的地理位置分发中列出的步骤执行操作。有关配置第三方服务的信息，请参阅使用第三方地理位置服务。

来源访问标识 (OAI)

通常，如果您使用 Amazon S3 存储桶作为 Amazon CloudFront 的来源，那么您需要向每个人授予读取您存储桶中对象的权限。这使得任何人都可使用 Amazon CloudFront URL 或 Amazon S3 URL 访问您的内容。Amazon CloudFront 不会暴露 Amazon S3 URL，但是，如果您的应用程序直接从 Amazon S3 提供任意内容或有任何人透露直接链接，攻击者就有可能发现这些 URL。

您可以通过创建 OAI（一种专用的 Amazon CloudFront 用户）来限制对 Amazon S3 的访问。您可以通过更改 Amazon S3 权限来向 Amazon CloudFront 赋予 OAI 权限，并移除其他所有权限。当您的最终用户使用 Amazon CloudFront 访问您的 Amazon S3 对象时，OAI 将代他们获取这些对象。如果最终用户尝试使用 Amazon S3 URL 访问对象，他们会被拒绝访问。

步骤 1：创建 CloudFront OAI 并将其添加到您的分配

步骤 2：授予 OAI 权限以读取 Amazon S3 存储桶中的对象

步骤 3：编辑 S3 存储桶权限

Amazon Route 53

Amazon Route 53 拥有两项使您更容易扩展基础设施和响应 DDoS 攻击的功能。它们是*别名记录集*和*私有 DNS*。

别名记录集

不同于普通的 Amazon Route 53 资源记录集，别名记录集提供的是 Amazon Route 53 – DNS 功能的特定扩展。别名记录集包含指向 Amazon CloudFront 分配、ELB 负载均衡器、Amazon S3 存储桶或同一托管区域中的另一 Amazon Route 53 资源记录集的指针，而非指向 IP 地址或域名。

别名记录集可以节省您的时间，并在遭受攻击时提供额外的工具。例如，假设 example.com 的别名记录集指向的是 ELB 负载均衡器，后者负责将流量分配到运行您应用程序的多个 EC2 实例。如果您的应用程序遭到攻击，则您可以更改别名记录集，使之指向 Amazon CloudFront 分配或具有运行 WAF 的更高 EC2 实例容量的其他 ELB 负载均衡器或您自己的安全工具。之后，Amazon Route 53 会在 example.com 的 DNS 响应中自动反映这些更改，而不对包含 example.com 别名记录集的托管区域做出任何更改。

创建别名记录集可提供灵活性：通过平时用不上但在遭受攻击时很有用的额外资源重定向流量。有关别名记录集的更多信息，请参阅选择别名资源记录集还是非别名资源记录集。

要创建别名记录集，请参阅使用 Amazon Route 53 创建别名记录集。

私有 DNS

私有 DNS 让您能够管理应用程序资源的内部 DNS 名称（Web 服务器、应用程序服务器、数据库等）,同时又不向公共 Internet 暴露此类信息。例如，如果您需要使用 CNAME 路由的内部资源，但不需要向外界公开它们，则可以使用 VPC 中的私有 DNS。

您也可以使用 Amazon Route 53 配置分割视图 DNS，又被称作水平分割 DNS。如果您想维持同一网站或应用程序的内部和外部版本（例如，分离管理和最终用户流量），则您可以配置公有和私有托管区域,从而针对同一域名返回不同的内部和外部 IP 地址。您只需要创建拥有同一域名的公有托管区域和私有托管区域，并在两个托管区域中创建相同的子域即可。

步骤 1：创建私有托管区域

步骤 2：列出私有托管区域

步骤 3：将 Amazon VPC 关联到私有托管区域

Web 应用程序防火墙

与针对基础设施的攻击相比，发生在应用层的 DDoS 攻击通常针对的是具有较低流量的 Web 应用程序。为缓解此类攻击，您将需要配置 WAF 作为您基础设施的一部分。

WAF 充当过滤器的角色，负责将一组规则应用到 Web 流量。通常，这些规则可弥补跨站点脚本 (XSS) 和 SQL 注入 (SQLi) 等漏洞，而且还可通过缓解 HTTP GET 或 POST 泛洪来构建抵御 DDoS 的弹性。

HTTP 充当最终用户与应用程序之间的请求-响应协议，其中，最终用户负责请求数据 (GET) 和提交待处理的数据 (POST)。GET 泛洪通过以较高的速率请求同一个 URL 或请求您应用程序的所有对象来实施攻击。POST 泛洪的攻击原理则是：寻找代价高昂的应用程序处理过程（例如登录或数据库搜索等操作），触发这些过程以“淹没”您的应用程序。

WAF 具有多项功能，都可以帮助避免此类攻击影响到您应用程序的可用性。其中之一是 HTTP 速率限制功能，它可让您设置受支持的 HTTP 请求阈值，以限制每位最终用户在特定时间段内能够发起的请求数量。一旦最终用户的请求数量超出了该阈值，WAF 可以阻止或缓冲新的请求，以确保其他最终用户仍可访问您的应用程序。

此外，WAF 还能检查 HTTP 的请求并识别不遵守正常模式的请求，如：阻止超出字符数限制的登录、阻止搜索所有项目的操作等。其他有助于抵御应用层 DDoS 攻击的 WAF 功能有“全自动区分计算机和人类的图灵测试”(CAPTCHA) 以及 IP 声誉列表。

要在您的 AWS 基础设施中部署 WAF，首先，您需要在 AWS Marketplace 中选择一种可用的 WAF 解决方案。AWS Marketplace 是一个在线商店，可帮助 AWS 客户查找、购买和安装可在 EC2 实例上运行的软件。您可在 AWS Marketplace 安全类别下或通过搜索 Web 应用程序防火墙来找到 WAF 解决方案。

选定 WAF 解决方案后，您需要将该软件部署到 EC2 实例上并配置实例，让其能够随您的流量扩展。在此之前，我们先来讨论一下实现 AWS Marketplace WAF 扩展的其他要求。

要检查所有的 HTTP 请求，WAF 必须位于您应用程序的流量通路上。不幸的是，这会造成 WAF 成为故障点或瓶颈的情况。为缓解这一问题，您需要在流量峰值期间根据需要有运行多个 WAF 的能力。您可通过“WAF 三明治”来实现这种类型的 WAF 扩展。

在“WAF 三明治”中，运行 WAF 软件的 EC2 实例包含在 Auto Scaling 组中，并介于两个 ELB 负载均衡器之间。回想在 Elastic Load Balancing 章节中，您创建了两个负载均衡器：默认 VPC 中的一个基本负载均衡器，以及一个内部负载均衡器。默认 VPC 中的基本负载均衡器将作为前端，该负载均衡器面向公共网络，负责将所有入站流量分配到 WAF EC2 实例中。通过在 ELB 后的 Auto Scaling 组中运行 WAF EC2 实例，该实例可在流量峰值达到较高级别时扩展并增加 WAF EC2 实例数。

一旦流量被检查并过滤后，WAF EC2 实例将其转发到内部的后端负载均衡器，后者再将流量分配到您的应用程序 EC2 实例。该配置（如图 5 所示）能够让 WAF EC2 实例在不影响应用程序 EC2 实例可用性的前提下扩展并满足容量需求。

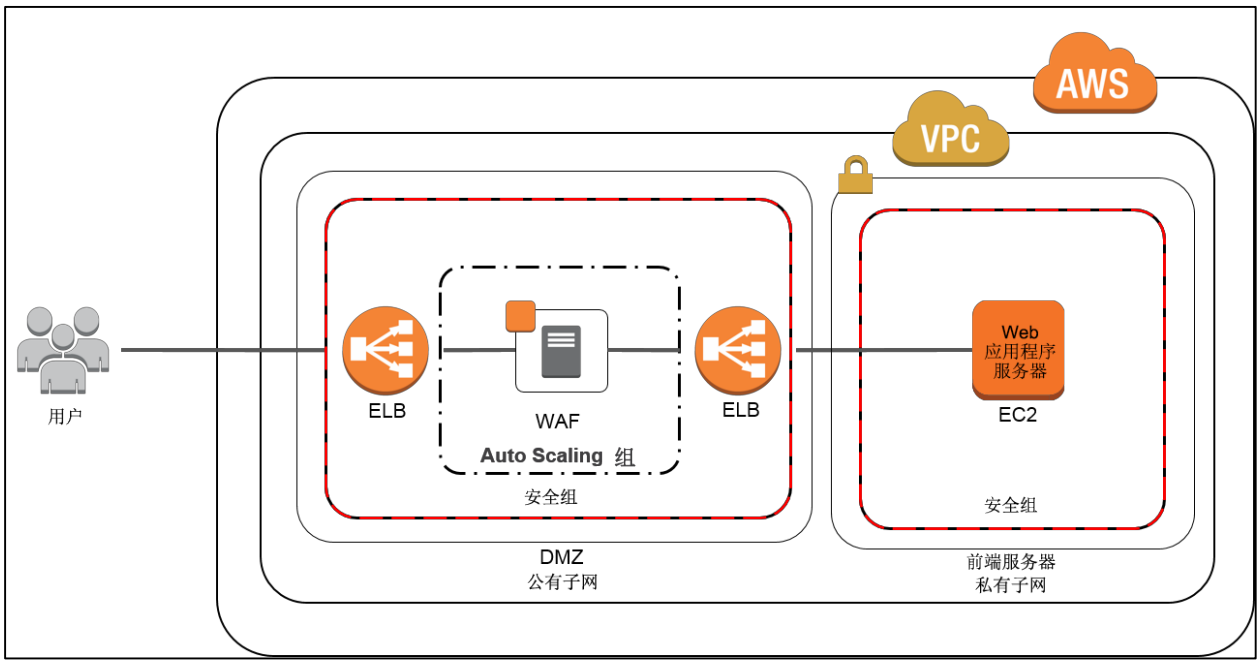

图 5：“WAF 三明治”

由于我们已经创建了基本和内部 ELB 负载平衡器，因此，下一步是在两个 ELB 负载均衡器之间的 EC2 实例上部署并配置 WAF。AWS Marketplace 解决方案采用不同的配置来支持“WAF 三明治”的 Auto Scaling。从 AWS Marketplace 选择某个 WAF 后，您应联系 AWS 合作伙伴，以索要文档或要求其协助您部署和扩展软件。此外，您还可以联系 AWS 专业服务来帮助您在基础设施中部署 WAF。

了解正常行为

确保实现弹性架构的途径之一是知道您的应用程序或基础设施何时会受到攻击。通常，客户会通过其应用程序是否可用来判定攻击状态。但是，更好的策略是了解您的应用程序预期的流量级别和模式，并以此为基准来识别异常的级别或模式。

DDoS 攻击者通常会通过确定阈值的方式来探测或测试您的应用程序。一旦完成几次试探性攻击，攻击者就能估算出可用于成功影响到您的应用程序可用性的流量或向量水平。在这些情况下，知道预期结果和非预期结果可帮助您识别异常并得到预警，以帮助建立态势感知能力。

Amazon CloudWatch

您可以使用 Amazon CloudWatch 监控您在 AWS 中运行的基础设施和应用程序。Amazon CloudWatch 能够收集指标、日志文件，并设置在这些指标超出预定阈值时发出警报。此外，如果您不了解正常情况应当如何，则您还可以使用 Amazon CloudWatch 全面系统地了解资源使用率、应用程序性能和运行状况。您可以借助这些信息来应对 DDoS 攻击，并评估何时需要进行更改。

在创建警报前，您需要熟悉应用程序的正常执行情况。为此，您可以查看、选择和绘制 Amazon CloudWatch 指标。

步骤 1：查看可用指标

步骤 2：搜索可用指标

步骤 3：选择指标

步骤 4：获取指标的统计数据

步骤 5：绘制指标图表

Amazon CloudFront 报告和 Amazon Route 53 健康检查都有单独的指标，您可以通过查看来更好的了解您的流量和应用程序。

一旦为您的应用程序设立了基准，您就可以借助 Amazon CloudWatch 创建基于这些指标的警报，以便在出现可能的攻击行为或其他违规行为时提醒您注意。下表包含了建议用于 DDoS 攻击的警报指标。

主题	指标	描述
Auto Scaling	GroupMaxSize	Auto Scaling 组的最大规模。
AWS 账单	EstimatedCharges	针对您的 AWS 使用率的预计收费。
Amazon CloudFront	Request	所有 HTTP/S 请求的数目。

Amazon CloudFront	TotalErrorRate	HTTP 状态代码为 4xx 或 5xx 的所有请求所占的百分比。
Amazon EC2	CPUUtilization	当前正在使用的已分配 EC2 计算单位的百分率?。
Amazon EC2	NetworkIn	实例在所有网络接口上收到的字节数。
Amazon EC2	StatusCheckFailed	报告以下两种状况检查之一是否失败：StatusCheckFailed_Instance 或 StatusCheckFailed_System 。
ELB	UnHealthyHostCount	每个可用区中运行状况不佳的实例的数量。
ELB	RequestCount	已接收并路由到注册实例且已完成请求的数量。
ELB	Latency	请求离开负载均衡器直至收到响应所用的时间（以秒为单位）。
ELB	HTTPCode_ELB_4xx HTTPCode_ELB_5xx	负载均衡器生成的 HTTP 4XX 或 5XX 错误代码的数量。
ELB	BackendConnectionErrors	失败连接的数量。
ELB	SpilloverCount	因队列已满而被拒绝的请求的数量。
Amazon Route 53	HealthCheckStatus	终端节点的健康检查状态。

表 1：推荐的 CloudWatch 指标

除了这些默认的指标外，您还可以使用 Amazon CloudWatch Logs 来监控和访问在 EC2 实例上运行的应用程序的日志文件。Amazon CloudWatch Logs 是一种可在 Amazon Linux、Ubuntu 和 Windows 上安装的代理，它负责将日志发送到 Amazon CloudWatch。Amazon CloudWatch Logs 能够跟踪应用程序日志中的错误数，并在错误率超过指定阈值时向您发送通知。

此外，您还可以监控应用程序日志以查找特定字词（如“NullReferenceException”）或日志数据中特定位置处的某个字词（如 Apache 访问日志中的“404”状态代码）出现的次数。找到所需的字词后，Amazon CloudWatch Logs 会向您指定的 Amazon CloudWatch 指标报告该数据。

本章节讲解在运行 Amazon Linux 或 Ubuntu 服务器的 Amazon EC2 实例中安装 Amazon CloudWatch Logs 代理的步骤。要在运行 Windows 的 Amazon EC2 实例上开始使用 Amazon CloudWatch Logs，请参阅*适用于 Microsoft Windows 实例的 Amazon EC2 用户指南* 中的将性能计数器发送到 CloudWatch 和将日志发送到 CloudWatch Logs。

步骤 1：快速入门：在现有 EC2 实例中安装和配置 CloudWatch Logs 代理

步骤 2：设置 Amazon Simple Notification Service

步骤 3：创建警报

有关可用警报的更多信息，请参阅 Amazon CloudWatch 命名空间、维度和指标参考。

制订攻击防范计划

在受到攻击时才制订的应对策略往往不会奏效。应在受到攻击前将计划部署到位，以确保：

- 您已验证架构并选择了适用于您的基础设施和应用程序的技术。
- 您已评估了提高弹性的成本并明白您的防御目标。
- 当受到攻击时，您知道该联系谁。

在制订该计划时，您应考虑自己需要的支持级别。AWS 为所有客户提供高度个性化的支持。但是，在 DDoS 攻击期间寻求获得更高级别支持的客户则应考虑企业级支持。企业级支持具有以下好处：

- 技术客户经理 (TAM)：TAM 具有 AWS 全部服务的专业技术知识，并且详细了解您的架构。TAM 与 AWS 解决方案架构师进行协作来帮助您遵循最佳实践，并可与主要联络人直接通话，以满足您的持续支持需求。
- 服务周到的优质案例转交：案例将直接转给经过专门培训的工程师以帮助确保快速、准确地解决关键问题。

有关此类支持功能和不同支持级别的更多信息，请访问 AWS Support 中心。

总结

AWS 提供了大量可供您提高弹性以抵御 DDoS 攻击的服务和功能。您负责适当地使用这些以及其他措施，以保护您在云中的基础设施和应用程序，满足您对于 DDoS 防护的要求。AWS 鼓励您运用本文中介绍的最佳实践，来制订一组安全策略和应对措施，从而帮助您提高抵御 DDoS 攻击的弹性。