实现 DDoS 防御能力的 AWS 最佳实践

2016 年 6 月

版权声明

目录

目录 3
摘要 4
介绍 4
DDoS 攻击 4
基础设施层攻击 6
应用层攻击 7
缓解技术 8
基础设施层防御 (BP1、BP3、BP6、BP7) 11
应用层防御 (BP1、BP2、BP6) 13
缩小攻击面 15
模糊 AWS 资源 (BP1、BP4、BP5) 16
操作技术 18
可见性 18
支持 20
结论 21
贡献者 21
备注 21

摘要

本文面向想要改进其在 Amazon Web Services (AWS) 上运行的应用程序的抗风险能力以抵御分布式拒绝服务 (DDoS) 攻击的客户。本文概述了 DDoS 攻击、AWS 提供的功能、缓解技术以及可用作有助于保护应用程序可用性的指南的 DDoS 防御参考架构。

介绍

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。每一节都有指向 AWS 文档的链接，这些文档提供了有关最佳实践或功能的更多详细信息。您还可以查看 AWS re:Invent 会议场次 SEC307 — 使用 AWS 构建 DDoS 防御架构[1]和 SEC306 — 防御 DDoS 攻击[2]，以了解更多信息。

DDoS 攻击

拒绝服务 (DoS) 攻击是指可能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从单一来源对目标实施攻击，如图 1 所示。

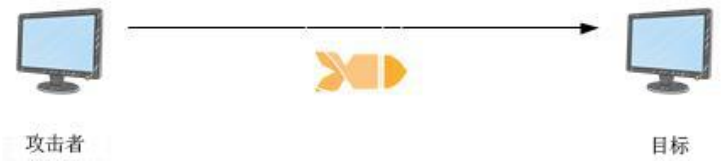

图 1: DoS 攻击示意图

在分布式拒绝服务 (DDoS) 攻击形式中，攻击者将借助多种来源 (可能遭到一组协作者的盗用或控制) 发动对目标的攻击。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。

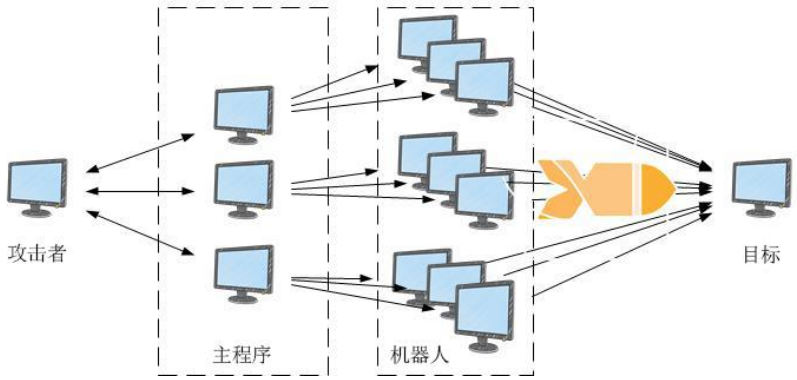

图 2: DDOS 攻击示意图

DDoS 攻击最常发生在开放系统互连 (OSI) 模型的第 3、4、6 和 7 层，如表 1 中所述。第 3 和第 4 层攻击对应于 OSI 模型的网络层和传输层，本文档将它们称为基础设施层攻击。第 6 和第 7 层攻击对应于 OSI 模型的展示层和应用层，本文档将它们称为应用层攻击。

#	层	单位	描述	媒介示例
7	应用程序	数据	应用程序的网络进程	HTTP 泛洪、DNS 查询泛洪
6	展示	数据	数据表示和加密	SSL 滥用
5	*会话*	*数据*	*主机间通信*	*不适用*
4	传输	分段	端到端连接和可靠性	SYN 泛洪
3	网络	数据包	路径确定和逻辑寻址	UDP 反射攻击
2	*数据链路*	*帧*	*物理寻址*	*不适用*
1	*物理*	*Bits*	*媒体、信号和二进制传输*	*不适用*

表 1: 开放系统互连 (OSI) 模型

这种区分非常重要，因为针对这些层的攻击类型是不同的，因此要使用不同的技术来建立防御能力。

基础设施层攻击

常见的 DDoS 攻击、用户数据报协议 (UDP) 反射攻击和同步 (SYN) 泛洪都属于基础设施层攻击。攻击者可以使用这些方法之一来生成大量流量，从而使网络或系统(如服务器、防火墙、IPS 或负载均衡器) 的容量不胜负荷。这些攻击有明确的特征，很容易检测出来。有效缓解这些攻击需要超过攻击者生成的流量的网络或系统资源。

UDP 是无状态协议。这让攻击者能够哄骗发送到服务器的请求的源，以引出更大的响应。放大系数 (请求大小与响应大小的比值) 随所用的协议 (如域名系统 (DNS)、网络时间协议 (NTP) 或简单服务发现协议 (SSDP)) 而变。例如，DNS 的放大系数介于 28 到 54 之间，也就是说，攻击者向 DNS 服务器发送 64 字节的请求有效负载可以生成超过 3400 字节的非必要流量。图 3 中阐释了此概念。

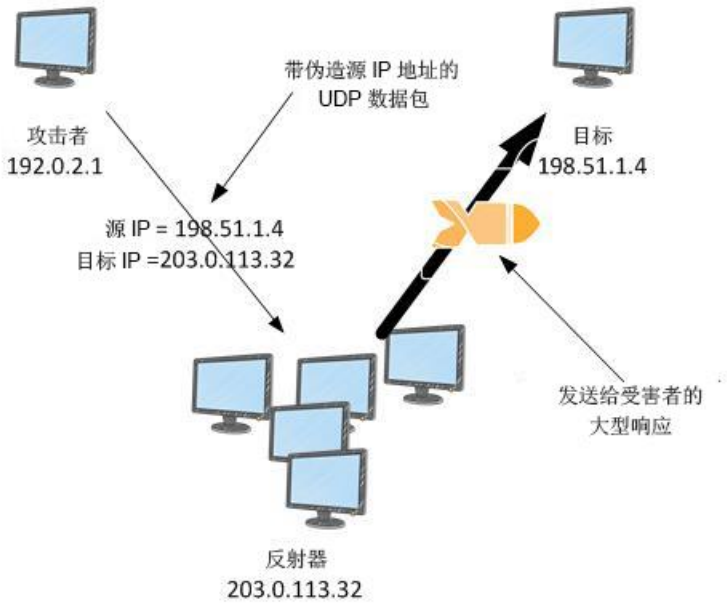

图 3: UDP 反射攻击

SYN 泛洪可以达到大约每秒数十 Gb 的速度，但攻击的目的是通过将连接保持在半开放状态来耗尽系统的可用资源。如图 4 所示，当最终用户连接到 TCP 服务 (如 Web 服务器) 时，客户端将发送 SYN 数据包。服务器将返回 SYN-ACK，客户端将返回 ACK，完成三向握手。

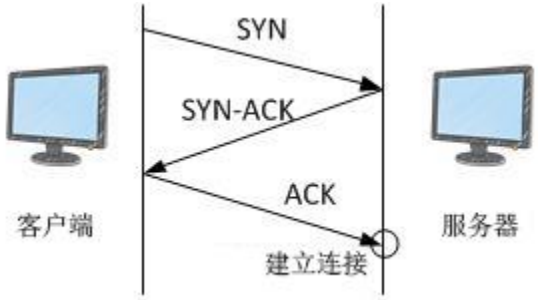

图 4: SYN 三向握手

在 SYN 泛洪中，永远不会返回 ACK，服务器将一直等待响应。这会使新用户无法连接到服务器。

应用层攻击

偶尔，攻击者可能使用第 7 层或应用层攻击来攻击应用程序本身。这些攻击与基础设施层攻击不同，因为攻击者试图过度行使应用程序的特定功能，以使其不可用。有些情况下，这可以通过不产生大量网络流量的极少请求量来实现。这可能使攻击更难以检测和缓解。应用层攻击的示例包括 HTTP 泛洪、缓存清除攻击和 WordPress XML-RPC 泛洪。

使用 HTTP 泛洪，攻击者发送看似来自 Web 应用程序的真实用户的 HTTP 请求。有些 HTTP 泛洪针对特定资源，而更复杂的 HTTP 泛洪将尝试模仿人类行为。这可能会增加诸如请求速率限制等常见缓解技术的使用难度。缓存清除攻击是一种 HTTP 泛洪，它在查询字符串中使用变体来规避内容分发网络 (CDN) 缓存，这会导致来源获取操作，从而给源 Web 服务器造成额外的压力。

使用 WordPress XML-RPC 泛洪 (也称为 WordPress pingback 泛洪)，攻击者可能滥用 WordPress 品牌内容管理软件上托管的网站的 XML-RPC API 函数来生成大量 HTTP 请求。pingback 功能允许在 WordPress 上托管的网站 (站点 A) 通知另一个 WordPress 站点 (站点 B)，指出站点 A 已经创建了到站点 B 的链接。因此，站点 B 将尝试获取站点 A 以验证链接是否存在。在出现 pingback 泛洪时，攻击者滥用此功能来导致站点 B 攻击站点 A。此类型的攻击具有明确的特征，因为 HTTP 请求标头的“User-Agent”中应存在“WordPress”字样。

应用层攻击还可以针对域名系统 (DNS) 服务。这些攻击中最常见的是 DNS 查询泛洪，其中攻击者使用许多格式正确的 DNS 查询来耗尽 DNS 服务器的资源。这些攻击还可以包括缓存清除部分，其中攻击者将子域字符串随机化以绕过任何给定解析器的本地 DNS 缓存。因此，解析器不知不觉地被用来发起针对权威 DNS 服务器的攻击。

对于通过安全套接字层 (SSL) 交付的 Web 应用程序，攻击者可以选择攻击 SSL 协商过程。SSL 需要进行大量的计算，这使得攻击者可以通过发送难以理解的数据来影响服务器的可用性。这种攻击的其他变体包括攻击者虽然完成 SSL 握手，但不断地重新协商加密方法。类似地，攻击者可以选择通过打开和关闭许多 SSL 会话来耗尽服务器资源。

缓解技术

AWS 基础设施采用能够防御 DDoS 的设计，并由 DDoS 缓解系统提供支持，这些系统可以自动检测和过滤多余流量。为了保护应用程序的可用性，有必要实施一种允许您利用这些功能的架构。

最常见的 AWS 使用案例之一是通过 Internet 向用户提供静态和动态内容的 Web 应用程序。有关通常与 Web 应用程序结合使用的 DDoS 防御参考架构，请参阅图 5。

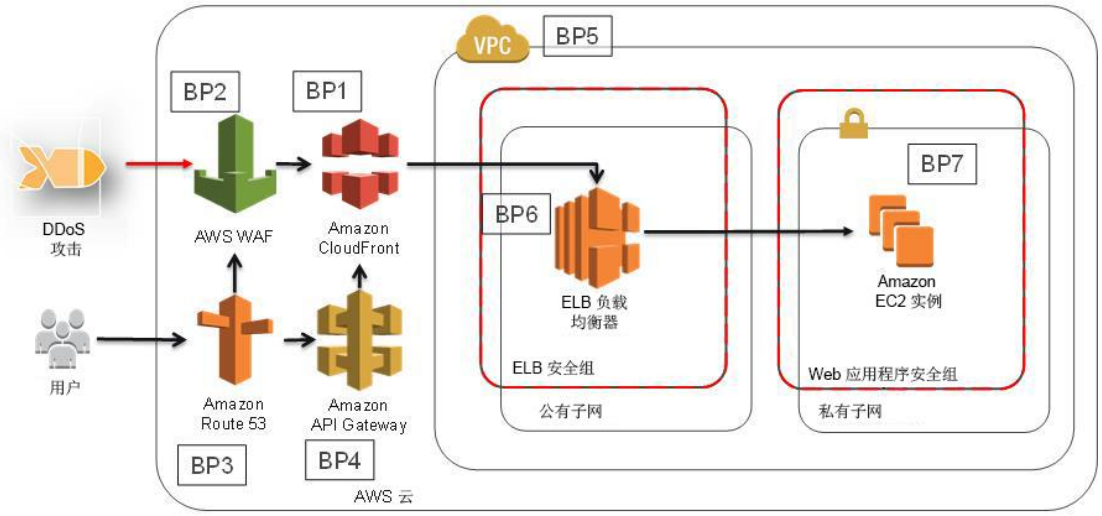

图 5: DDoS 防御参考架构

此参考架构包括许多 AWS 服务，可帮助您提高 Web 应用程序防御 DDoS 攻击的能力。我们列出了此架构中的最佳实践，以便在整个文档中讨论时进行参考。例如，讨论 Amazon CloudFront 提供的功能的一节将使用最佳实践指示器 (如 BP1) 加以引用。有关这些服务及其提供的功能的汇总，请参阅表 2。

	AWS 边缘站点			**AWS 区域**		
	Amazon CloudFront + AWS WAF (BP1、BP2)	Amazon API Gateway (BP4)	Amazon Route 53 (BP3)	Elastic Load Balancing (BP6)	Amazon VPC (BP5)	Amazon EC2 + Auto Scaling (BP7)
第 3 层 (如 UDP 反射) 攻击缓解	✔	✔	✔	✔	✔	
第 4 层是 (如 SYN 泛洪) 攻击缓解	✔	✔	✔	✔		
第 6 层 (如 SSL) 攻击缓解	✔	✔	不适用	✔		
缩小攻击面	✔	✔	✔	✔	✔	
扩展以吸收应用层流量	✔	✔	✔	✔		✔
第 7 层 (应用层) 攻击缓解	✔	✔	✔			
地理隔离以及分散多余流量和更大的 DDoS 攻击	✔	✔	✔			

表 2: 最佳实践汇总

AWS 区域中可用的服务 (如 Elastic Load Balancing 和 Amazon Elastic Compute Cloud (EC2)) 允许您打造 DDoS 防御能力和规模以处理给定区域内的意外流量。AWS 边缘站点中可用的服务 (如 Amazon CloudFront、AWS WAF、Amazon Route 53 和 Amazon API Gateway) 允许您利用全球边缘站点网络，为您的应用程序提供更大容错能力和更大规模来管理更多流量。使用其中每个服务为基础设施层和应用层建立抗风险能力以抵御 DDoS 攻击的好处将在以下各节中讨论。

基础设施层防御 (BP1、BP3、BP6、BP7)

在传统的数据中心环境中，您可以通过使用诸如过度配置容量、部署 DDoS 缓解系统或在 DDoS 缓解服务的帮助下清理流量等技术来缓解基础设施层 DDoS 攻击。在 AWS 上，您可以选择对应用程序的架构进行设计，以便能够扩展和吸收更多流量，而无需进行巨额投资，也不会产生不必要的复杂性。缓解容量耗尽型 DDoS 攻击的主要考虑因素包括传输能力的可用性和多样性，以及保护 Amazon EC2 实例等 AWS 资源免受攻击流量的影响。

实例大小 (BP7)

许多 AWS 客户使用 Amazon EC2 来获得大小可调的计算容量，这允许您在需求变化时快速扩展或收缩。需要时，您可以通过向应用程序添加实例来横向扩展。还可以选择使用较大的实例进行纵向扩展。有些实例类型支持 10 Gb 网络接口和增强联网等功能，这些功能可以提高您处理更多流量的能力。

使用 10 Gb 网络接口，每个实例都能够支持更多流量。这有助于防止已到达 Amazon EC2 实例的任何流量出现接口拥塞。与传统实施相比，支持增强联网的实例可提供更高的 I/O 性能和更低的 CPU 利用率。这提高了实例处理具有较多数据包的流量的能力。在 AWS 上，您不负责入站数据传输的成本。

要了解有关支持 10 Gb 网络接口和增强联网的 Amazon EC2 实例的更多信息，请参阅 Amazon EC2 实例类型[3]。要了解如何启用增强联网，请参阅在 VPC 中对 Linux 实例启用增强联网[4]。

区域选择 (BP7)

许多 AWS 服务 (如 Amazon EC2) 都可在全球多个位置使用。这些在地理上分开的区域称为 AWS 区域。在设计应用程序的架构时，您可以根据自己的要求选择一个或多个区域。常见考虑事项包括性能、成本和数据主权。在每个区域中，AWS 都提供对一组唯一的 Internet 连接和对等关系的访问权限，以便对处境类似的最终用户提供较优延迟和吞吐量。

还必须结合 DDoS 防御能力考虑您对区域的选择。许多区域距离大型 Internet 交换点都比较近。由于许多 DDoS 攻击源自世界各地，因此，靠近交换点非常有帮助，国际运营商和大型同行在这些地方通常具有较强的实力。这有助于最终用户在处理更多流量时访问您的应用程序。

要了解有关选择区域的更多信息，请参阅区域和可用区[5]，并向您的客户团队了解每个区域的特征以帮助您作出明智的决策。

负载均衡 (BP6)

较大的 DDoS 攻击可能会超过单个 Amazon EC2 实例的大小。为了缓解这些攻击，您需要考虑对多余流量进行负载均衡的选项。使用 Elastic Load Balancing (ELB)，您可以通过在许多后端实例之间分发流量来降低应用程序过载的风险。ELB 可以自动扩展，允许您管理更多的意外流量，例如访问高峰或 DDoS 攻击。

ELB 只接受格式正确的 TCP 连接。这意味着许多常见的 DDoS 攻击 (如 SYN 泛洪或 UDP 反射攻击) 将不被 ELB 接受，并且不会传递到您的应用程序。当 ELB 检测到这些类型的攻击时，它会自动扩展以吸收额外的流量，但不会产生任何额外费用。

要了解有关使用 ELB 分发负载和保护 Amazon EC2 实例的更多信息，请参阅 Elastic Load Balancing 入门[6]。

使用 AWS 边缘站点大规模交付 (BP1、BP3)

访问高度可扩展的多样化 Internet 连接可显著提高您为最终用户优化延迟和吞吐量，吸收 DDoS 攻击以及隔离故障，同时最大限度降低可用性影响的能力。AWS 边缘站点提供了一个额外的网络基础设施层，通过使用 Amazon CloudFront 和 Amazon Route 53 为 Web 应用程序提供这些好处。通过这些服务，将从与您的最终用户距离更近的位置提供您的内容和解析 DNS 查询。

边缘的 Web 应用程序交付 (BP1)

Amazon CloudFront 是一种内容分发网络 (CDN) 服务，可用于交付您的整个网站，包括静态、动态、流媒体和交互内容。持久 TCP 连接和可变生存时间 (TTL) 可用于加快内容交付速度，即使内容不能在边缘站点缓存也是如此。这允许您使用 Amazon CloudFront 来保护您的 Web 应用程序，即使您不提供静态内容。Amazon CloudFront 仅接受格式正确的连接，以防止许多常见的 DDoS 攻击 (如 SYN 泛洪和 UDP 反射攻击) 到达您的原始服务器。DDoS 攻击在地理上与原始服务器相隔离，这可防止流量影响其他位置。这些功能可以极大地提高您在大型 DDoS 攻击期间继续为最终用户提供流量的能力。您可以使用 Amazon CloudFront 来保护 AWS 或 Internet 上其他位置的原始服务器。

要了解有关使用 Amazon CloudFront 优化 Web 应用程序性能的更多信息，请参阅 CloudFront 入门[7]。

边缘的域名解析 (BP3)

Amazon Route 53 是一种高度可用且可扩展的域名系统 (DNS) 服务，可用于将流量定向到您的 Web 应用程序。它包括许多高级功能，例如流量、基于延迟的路由、地理 DNS、运行状况检查和监控。这些功能允许您控制服务如何响应 DNS 请求，以便针对延迟、运行状况和其他注意事项进行优化。您可以使用这些功能来提高 Web 应用程序的性能并避免站点中断。

Amazon Route 53 使用随机分区和任播条带化，以使最终用户也能够访问您的应用程序，即使 DNS 服务是 DDoS 攻击的目标也是如此。使用随机分区，您的委托集中的每个名称服务器都对应一组唯一的边缘站点和 Internet 路径。这可提供更强的容错能力并尽可能减少客户之间的重叠。如果委派集中的一个名称服务器不可用，则最终用户可以重试并接收其他边缘站点中的另一个名称服务器的响应。任播条带化的使用使得每个 DNS 请求都由最佳位置进行处理。这会产生分散负载并减少 DNS 延迟的效果，从而使最终用户能够更快地收到响应。此外，Amazon Route 53 还可以检测 DNS 查询的来源和数量异常，并优先处理来自已知可靠的用户的请求。

如果您有许多 Amazon Route 53 托管区域，则可以创建可重用的委派集，为每个域提供一组相同的权威名称服务器。这可以简化托管区域的维护。如果出现 DDoS 攻击，它还允许 AWS 应用单个缓解措施来覆盖所有使用可重用委派集的托管区域。

要了解有关使用 Amazon Route 53 将最终用户定向到您的应用程序的更多信息，请参阅 Amazon Route 53 入门[8]。要了解有关可重用委派集的更多信息，请参阅可重用委派集的操作[9]。

应用层防御 (BP1、BP2、BP6)

本文讨论的许多技术都可以有效地缓解基础设施层 DDoS 攻击对可用性的影响。保护应用程序免遭应用层攻击需要您实现一种架构，该架构允许您检测并扩展以吸收和阻止恶意请求。这是一个重要的考虑因素，因为基于网络的 DDoS 缓解系统在缓解复杂的应用层攻击方面通常无效。

检测和过滤恶意 Web 请求 (BP1、BP2)

Web 应用程序防火墙 (WAF) 通常用于保护 Web 应用程序以抵御试图利用应用程序中的漏洞的攻击。常见示例包括 SQL 注入或跨站点请求伪造。您还可以使用 WAF 来检测和缓解 Web 应用层 DDoS 攻击。

在 AWS 上，您可以使用 Amazon CloudFront 和 AWS WAF 保护您的应用程序免受这些攻击。Amazon CloudFront 允许您缓存静态内容，并从 AWS 边缘站点提供这些内容，从而帮助减轻原始服务器的负载。此外，Amazon CloudFront 还可以自动关闭来自慢速读取或慢速写入攻击者 (例如 Slowloris) 的连接。您可以使用 Amazon CloudFront 地理位置限制来阻止特定地理位置的用户访问您的内容。如果您想要阻止来自您不希望为最终用户提供服务的地理位置的攻击，这可能非常有用。

对于其他类型的攻击，如 HTTP 泛洪或 WordPress Pingback 泛洪，您可以使用 AWS WAF 创建自己的缓解措施。如果您知道要阻止的源 IP 地址，则可以创建一个具有阻止操作的规则，并将其与 Web ACL 相关联。然后，您可以在 Web ACL 中创建 IP 地址匹配条件，以阻止参与攻击的源 IP 地址。还可以使用按 URI、查询字符串、HTTP 方法或标头关键字进行阻止的条件创建规则。标头关键字在具有明确特征的攻击中非常有用。例如，WordPress pingback 攻击的 User-Agent 中将始终具有“WordPress”字样。

识别 DDoS 攻击的特征或准确识别参与攻击的 IP 地址可能颇具挑战性。有时，可以通过查看 Web 服务器日志来查找此信息。您还可以使用 AWS WAF 控制台查看 Amazon CloudFront 已转发给 AWS WAF 的请求示例。请求示例可以帮助确定可能需要哪些规则来缓解应用层攻击。如果您看到许多带有随机查询字符串的请求，可以决定在 Amazon CloudFront 中禁用查询字符串转发。这可以帮助缓解针对您的原始服务器的缓存清除攻击。

有些攻击由伪装成正常最终用户流量的 Web 流量组成。要缓解此类型的攻击，您可以使用 AWS Lambda 函数来实现基于速率的黑名单。使用基于速率的黑名单，您可以对 Web 应用程序可以处理的请求数量设置阈值。如果自动程序或爬网程序超过此限制，可以使用 AWS WAF 自动阻止任何额外请求。

要了解有关使用地理位置限制来限制对 Amazon CloudFront 分发的访问的更多信息，请参阅限制内容的地理位置分发[10]。

要了解有关使用 AWS WAF 的更多信息，请参阅 AWS WAF 入门[11]和查看 CloudFront 已转发到 AWS WAF 的 Web 请求的示例[12]。

要了解有关如何使用 AWS Lambda 和 AWS WAF 配置基于速率的黑名单的更多信息，请参阅如何使用 AWS WAF 和 AWS Lambda 配置基于速率的黑名单[13]。

扩展以吸收 (BP6)

处理应用层攻击的另一种方式是大规模操作。对于 Web 应用程序，您可以使用 ELB 将流量分配给许多为了处理流量激增而过度配置或配置为自动扩展的 Amazon EC2 实例，无论流量激增是访问高峰所致还是应用层 DDoS 攻击所致。Amazon CloudWatch 警报用于启动 Auto Scaling，该功能会根据您定义的事件自动调整 Amazon EC2 队列的大小。这保护了应用程序可用性，即使在处理意外请求数量时也是如此。通过使用 Amazon CloudFront 或 ELB，SSL 协商由分配或负载均衡器处理，这可以防止您的实例受到基于 SSL 的攻击的影响。

要了解有关使用 Amazon CloudWatch 调用 Auto Scaling 的更多信息，请参阅使用 Amazon CloudWatch 监控 Auto Scaling 实例和组[14]。

缩小攻击面

在 AWS 上进行架构设计时的另一个重要考虑因素是限制攻击者可以攻击您的应用程序的机会。例如，如果您不希望最终用户直接与某些资源进行交互，则需要确保这些资源不能从 Internet 访问。同样，如果您不希望最终用户或外部应用程序使用某些端口或协议与您的应用程序通信，则需要确保不接受流量。这个概念被称为缩小攻击面。在本节中，您将找到最佳做法，以缩小攻击面，并限制应用程序在 Internet 上的暴露程度。未暴露给 Internet 的资源更难以攻击，这限制了攻击者攻击应用程序可用性的选项。

模糊 AWS 资源 (BP1、BP4、BP5)

对于许多应用程序而言，您的 AWS 资源不需要完全暴露于 Internet。例如，ELB 后面的 Amazon EC2 实例可能不必可公开访问。在这种情况下，您可能决定允许最终用户在某些 TCP 端口上访问 ELB，并仅允许 ELB 与 Amazon EC2 实例通信。这可以通过在 Amazon Virtual Private Cloud (VPC) 中配置安全组和网络访问控制列表 (NACL) 来实现。Amazon VPC 允许您配置 AWS 云的逻辑隔离的部分，让您在自己定义的虚拟网络中启动 AWS 资源。

安全组和网络 ACL 的类似之处在于，它们都允许您控制对 VPC 内 AWS 资源的访问。安全组允许您在实例级别控制入站和出站流量，网络 ACL 提供类似的功能，但是在 VPC 子网级别提供。此外，Amazon EC2 安全组 (SG) 规则或网络 ACL 的入站数据传输不收取费用。这可确保您不需要为安全组或网络 ACL 丢弃的流量支付任何额外费用。

安全组 (BP5)

您可以在启动实例时指定安全组，或在稍后将实例与安全组相关联。除非创建*允许*规则以允许流量，否则从 Internet 到安全组的所有流量都会被隐式拒绝。例如，如果您有一个由 ELB 和许多 Amazon EC2 实例组成的 Web 应用程序，则可以决定为 ELB 创建一个安全组（“ELB 安全组”），为实例创建另一个安全组（“Web 应用程序服务器安全组”）。然后，您可以创建*允许*规则，允许从 Internet 到 ELB 安全组的流量，并允许从 ELB 安全组到 Web 应用程序服务器安全组的流量。因此，来自 Internet 的流量无法直接与 Amazon EC2 实例通信，这使得攻击者更难了解您的应用程序。

网络访问控制列表 (ACL) (BP5)

使用网络 ACL，您可以同时指定*允许*和*拒绝*规则。如果您希望明确拒绝应用程序的某些类型的流量，这非常有用。例如，您可以定义应对整个子网拒绝的 IP 地址(作为 CIDR 范围)、协议和目标端口。如果您的应用程序仅用于 TCP 流量，您可以创建规则来*拒绝*所有 UDP 流量，反之亦然。此工具在响应 DDoS 攻击时很有用，因为如果您知道源 IP 地址或其他特征，它可以允许您创建自己的规则来缓解攻击。

保护您的原始服务器 (BP1)

如果您正在将 Amazon CloudFront 与 VPC 内部的原始服务器结合使用，则应使用 AWS Lambda 函数自动更新安全组规则，以便仅*允许*来自 Amazon CloudFront 的流量。这可以通过帮助确保 Amazon CloudFront 和 AWS WAF 不会被绕过来提高您的原始服务器的安全性。

要了解有关通过自动更新安全组来保护原始服务器的更多信息，请参阅[如何通过使用 AWS Lambda 自动更新 Amazon CloudFront 和 AWS WAF 的安全组][15]。

您可能还需要确保只有 Amazon CloudFront 分配功能才能将请求转发到您的原始服务器。使用 Edge-to-Origin 请求标头，您可以在 Amazon CloudFront 将请求转发到原始服务器时添加标头或覆盖现有请求标头的值。您可以使用 *X-Shared-Secret* 标头来帮助验证您的原始服务器收到的请求由 Amazon CloudFront 发送。

要了解如何使用 *X-Shared-Secret* 标头来保护您的原始服务器的更多信息，请参阅[将自定义标头转发到您的原始服务器][16]。

保护 API 终端节点 (BP4)

通常，当需要向公众公开 API 时，存在 API 前端可能会成为 DDoS 攻击目标的风险。Amazon API Gateway 是一种完全托管的服务，允许您创建一个 API，作为在 Amazon EC2 或 AWS Lambda 上运行的应用程序或任何 Web 应用程序的“前门”。使用 Amazon API Gateway，您不需要为 API 前端运行自己的服务器，并且可以故意对公众模糊您应用程序的其他组件。这可以帮助防止这些 AWS 资源成为 DDoS 攻击的目标。Amazon API Gateway 与 Amazon CloudFront 集成，这使您能够受益于该服务固有的额外 DDoS 抵御能力。您还可以通过为 REST API 中的每个方法配置标准或突发速率限制来避免过多流量，从而保护后端。

要了解有关使用 Amazon API Gateway 创建 API 的更多信息，请参阅 [Amazon API Gateway 入门][17]。

操作技术

本文中的缓解技术允许您构建天生就能够抵御 DDoS 攻击的应用程序。许多情况下，了解 DDoS 攻击何时针对您的应用程序并能够对此数据采取措施也很有用。您还可能需要使用其他资源来评估威胁，查看应用程序的架构或请求其他帮助。本节讨论了以下方面的最佳实践：了解异常行为、警报和自动化以及接洽 AWS 以获得更多支持。

可见性

了解应用程序的正常行为让您在检测到异常时能够更快地采取行动。当关键指标严重偏离预期值时，表明攻击者可能正试图攻击您应用程序的可用性。使用 Amazon CloudWatch，您可以监控在 AWS 上运行的应用程序。它允许您收集和跟踪指标，收集和监控日志文件，设置警报，并自动应对您的 AWS 资源的变化。有关通常用来检测和应对 DDoS 攻击的 Amazon CloudWatch 指标的说明，请参阅表 3。

主题	指标	描述
Auto Scaling	GroupMaxSize	Auto Scaling 组的最大大小
Amazon CloudFront	Requests	HTTP/S 请求的数量
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 计算单位的百分比
Amazon EC2	NetworkIn	实例在所有网络接口上收到的字节数
ELB	SurgeQueueLength	由负载均衡器排队的请求数，等待后端实例接受连接并处理请求
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的已完成请求的数量
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	终端节点的运行状况检查状态

表 3: 推荐的 Amazon CloudWatch 指标

对于根据图 5 中提供的抗 DDoS 参考架构构建的应用程序，常见的基础设施层攻击将在到达应用程序之前被阻止。因此，这些攻击不会出现在您的 Amazon CloudWatch 指标中。

应用层攻击可能导致许多这些指标的升高。例如，HTTP 泛洪可能会导致 Amazon CloudFront、ELB 和 Amazon EC2 的请求数以及 CPU 和网络利用率指标升高。如果后端实例无法处理过多的请求，您还可能会看到 Amazon CloudFront 上的 TotalErrorRate 以及 ELB 上的 SurgeQueueLength、UnHealthyHostCount、Latency、BackendConnectionErrors、SpilloverCount 或 HTTPCode 升高。在这种情况下，HTTP 请求的量可能会被抑制，因为应用程序不能满足最终用户的正常需求。您可以通过扩展应用程序的后端或如本文前面所述通过使用 AWS WAF 阻止过多流量来补救此情况。

要了解有关使用 Amazon CloudWatch 检测针对您的应用程序的 DDoS 攻击的更多信息，请参阅 Amazon CloudWatch 入门[18]。

您可以用来了解应用程序攻击流量的另一个工具是 VPC 流日志。在传统网络上，您可以使用网络流日志来排查连接和安全问题，并确保网络访问规则正常工作。使用 VPC 流日志，您可以捕获有关在您的 VPC 中传入和传出网络接口的 IP 流量的信息。

每个流日志记录都包括源和目标 IP 地址、源和目标端口、协议，以及在捕获窗口期间传输的数据包数和字节数。该信息可用于帮助识别网络流量中的异常并识别特定的攻击媒介。例如，大多数 UDP 反射攻击将具有特定源端口 (例如，用于 DNS 反射的源端口 53)。这是一个明确的特征，您可以在流日志记录中轻松识别出来。作为回应，您可以选择在实例级别阻止特定源端口，或者创建网络 ACL 规则以阻止整个协议 (如果不需要该协议)。

要了解有关使用 VPC 流日志识别网络异常和 DDoS 攻击媒介的更多信息，请参阅 VPC 流日志[19]和 VPC 流日志 - 记录和查看网络流量流[20]。

支持

必须在 DDoS 攻击实际发生之前制定防御计划，这一点非常重要。本文中概述的最佳做法是采取积极措施，并且应在启动可能成为 DDoS 攻击目标的应用程序之前实施。您的客户团队可以帮助审核您的使用案例和应用程序，并帮助解决您可能遇到的任何具体问题或挑战:

有时，您会发现在 DDoS 攻击期间联系 AWS 以获得更多支持非常有益；您的案例将会迅速得到应答，并转交给能够提供帮助的专家。通过订阅商用级支持，您可以通过电子邮件、聊天或电话随时联系云支持工程师。

如果您在 AWS 上运行任务关键型工作负载，则应考虑企业级支持。使用企业级支持，您的紧急案例将会获得最高优先级，并转交给高级云支持工程师。此外，利用企业级支持，您还可以联系技术客户经理 (TAM)，此人是您的拥护者和专门的技术联系人。企业级支持还为您提供对基础设施事件管理服务的访问，该服务包括在计划的事件、产品发布和迁移期间的实时操作支持。

要了解有关选择支持计划以满足您的独特需求的更多信息，请参阅比较 AWS 支持计划[21]。

结论

本文中概述的最佳实践允许您构建一个能够保护应用程序的可用性，以抵御许多常见基础设施层和应用层 DDoS 攻击的抗 DDoS 架构。您能够根据这些最佳实践构建应用程序的程度将影响您能够缓解的 DDoS 攻击的类型、媒介和数量。AWS 鼓励您使用这些最佳实践来更好地保护应用程序的可用性，以抵御常见 DDoS 攻击。

贡献者

以下为对此文档有贡献的个人和组织：

- Andrew Kiggins，AWS 解决方案架构师
- Jeffrey Lyons，AWS DDoS 运营工程师

备注

[1] https://www.youtube.com/watch?v=OT2y3DzMEmQ

[2] https://www.youtube.com/watch?v=Ys0gG1koqJA

[3] https://aws.amazon.com/ec2/instance-types/

[4] http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/enhanced-networking.html

[5] http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-regions-availability-zones.html

[6] http://docs.aws.amazon.com/ElasticLoadBalancing/latest/DeveloperGuide/elb-getting-started.html

[7] http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/GettingStarted.html

[8] http://docs.aws.amazon.com/Route53/latest/DeveloperGuide/getting-started.html

[9] http://docs.aws.amazon.com/Route53/latest/APIReference/actions-on-reusable-delegation-sets.html

[10] http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/georestrictions.html

[11] http://docs.aws.amazon.com/waf/latest/developerguide/getting-started.html

[12] http://docs.aws.amazon.com/waf/latest/developerguide/web-acl-testing.html#web-acl-testing-view-sample

[13] https://blogs.aws.amazon.com/security/post/Tx1ZTM4DT0HRH0K/How-to-Configure-Rate-Based-Blacklisting-with-AWS-WAF-and-AWS-Lambda

[14] http://docs.aws.amazon.com/autoscaling/latest/userguide/as-instance-monitoring.html

[15] https://blogs.aws.amazon.com/security/post/Tx1LPI2H6Q6S5KC/How-to-Automatically-Update-Your-Security-Groups-for-Amazon-CloudFront-and-AWS-W

[16] http://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/forward-custom-headers.html

[17] https://aws.amazon.com/api-gateway/getting-started/

[18] http://docs.aws.amazon.com/AmazonCloudWatch/latest/DeveloperGuide/GettingStarted.html

[19] http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/flow-logs.html

[20] https://aws.amazon.com/blogs/aws/vpc-flow-logs-log-and-view-network-traffic-flows/

[21] https://aws.amazon.com/premiumsupport/compare-plans/