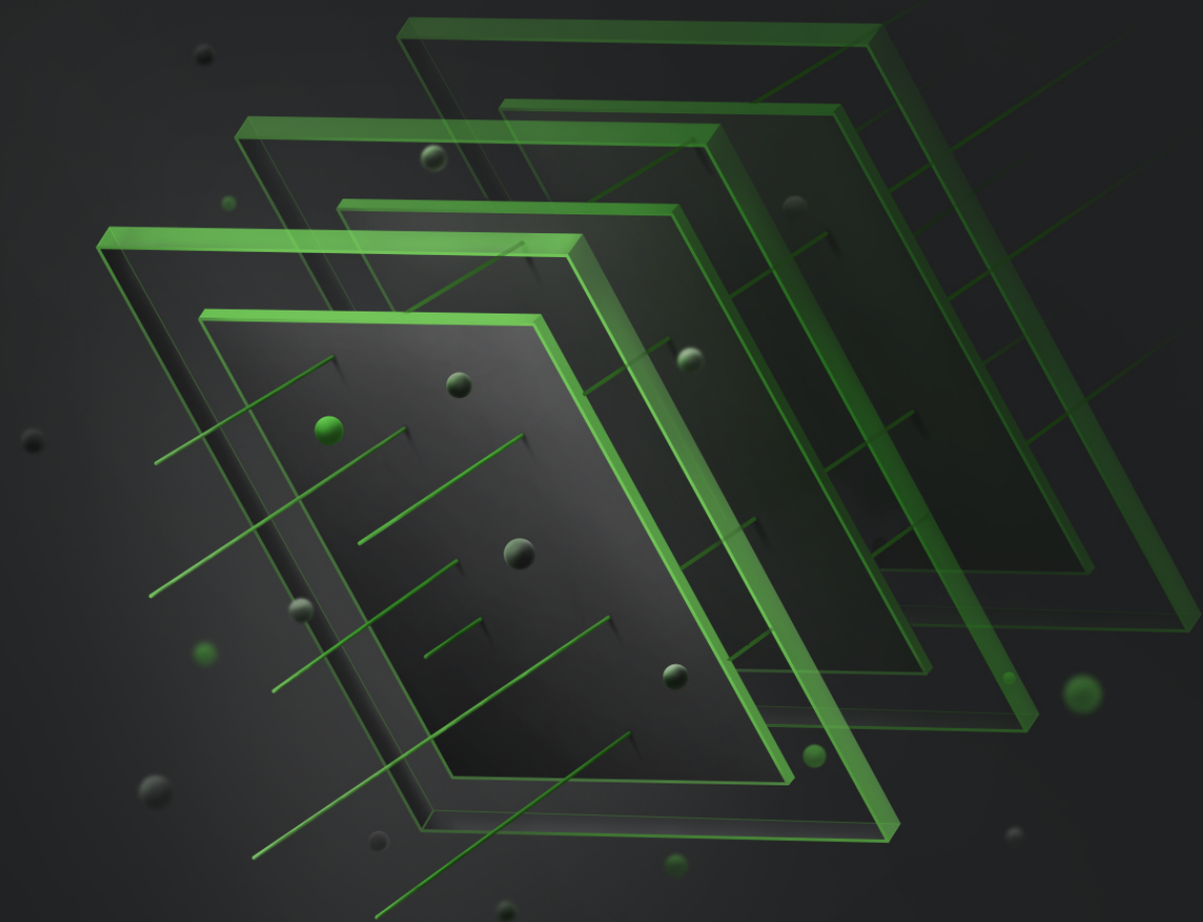
阿里云



2019年
上半年Web应用安全报告

阿里云安全团队

时间：2019.7

01

前言

PREFACE

Web应用安全依然是互联网安全的最大威胁来源之一，除了传统的网页和APP，API和各种小程序也作为新的流量入口快速崛起，更多的流量入口和更易用的调用方式在提高web应用开发效率的同时也带来了更多和更复杂的安全问题。一方面，传统的SQL注入、XSS、CC攻击等传统攻击手段和各种新爆出的web漏洞无时无刻不在考验着web应用安全方案的健壮性、灵活性和安全团队的快速反应能力，另一方面随着大数据技术和流量产业的成熟，互联网中来自自动化程序的流量占比也在迅速增长，爬虫也随之成为一个不容忽视的存在，伴随而来的数据泄露、流量作弊等问题也为各类业务带来了非常头痛的费用浪费、业务不可用以及各类业务安全类问题。

作为防御Web应用安全的基础设施，Web应用防火墙（WAF）依旧扮演着极其重要的角色，而其中云WAF又具备漏洞响应快、功能迭代迅速、支持弹性扩容、快速容灾等优势。本文根据阿里云WAF和防爬团队对2019年上半年云上流量的分析情况，为您带来最新的攻击趋势、漏洞应急情况以及一线安全专家的核心观点和防护建议。

02
攻击态势分析
ANALYSIS

1. 基础web攻击情况

传统web攻击手段中，WebShell上传/通信、SQL注入和命令执行依然是最常见的攻击行为，在传统攻击的分布上较往年没有太大的差异，如图2-1所示。

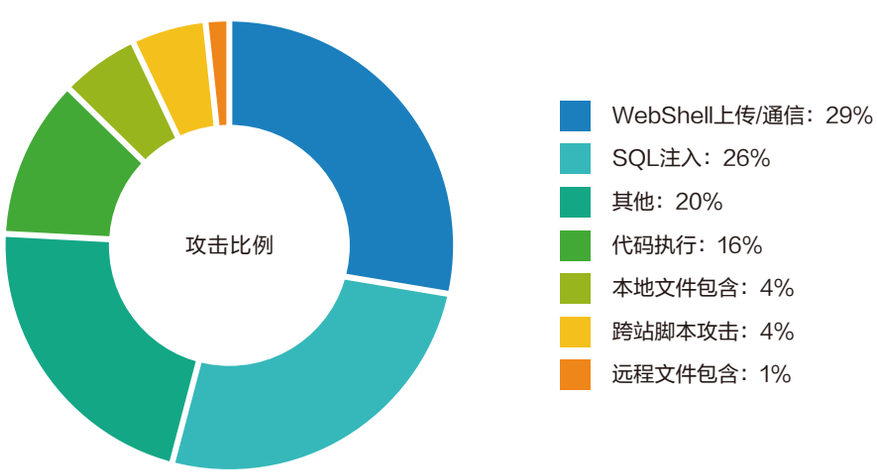


图2-1

从攻击源的分布来看，国内的攻击量仍然占据绝对优势，当然这与当前云上用户还是以国内用户和业务居多有关。除去国内的攻击源，美国和东南亚地区成为攻击源最多的地区，如图2-2所示。

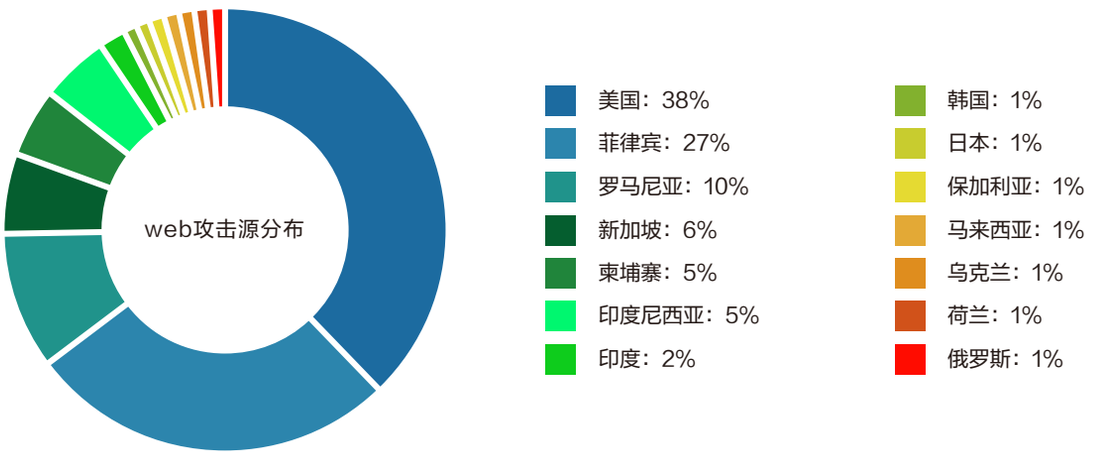


图2-2

从站点行业分布角度来看，备受攻击者青睐的网站多是互联网、电商、媒体和政府类站点，具体分布见下图：

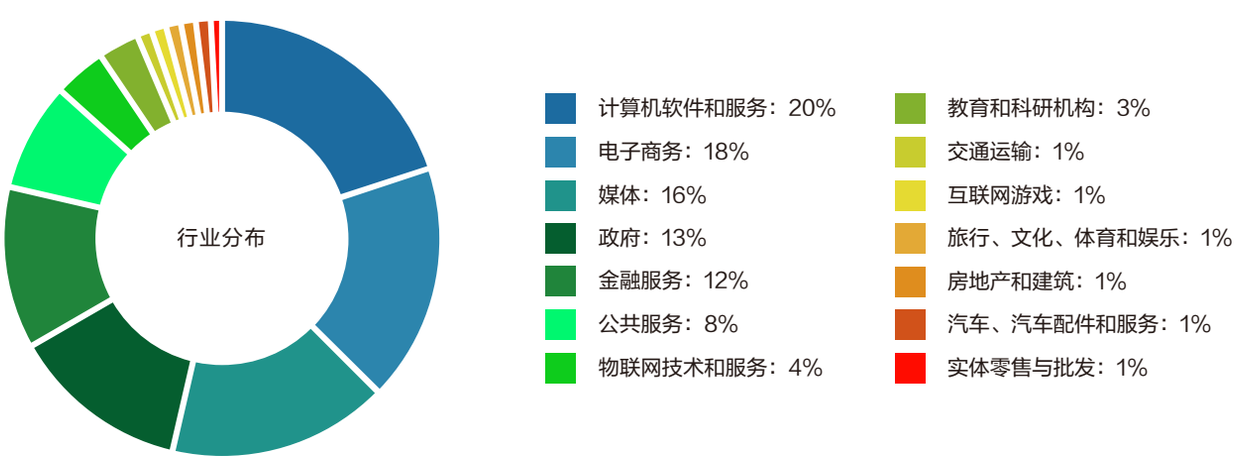


图2-3

从攻击的时间趋势来看，2019年上半年除了2月份春节以外，每个月的攻击次数都成递增趋势，到5月每个月拦截的攻击超过19亿，6月份拦截的攻击突破20亿，如下图所示。

攻击量

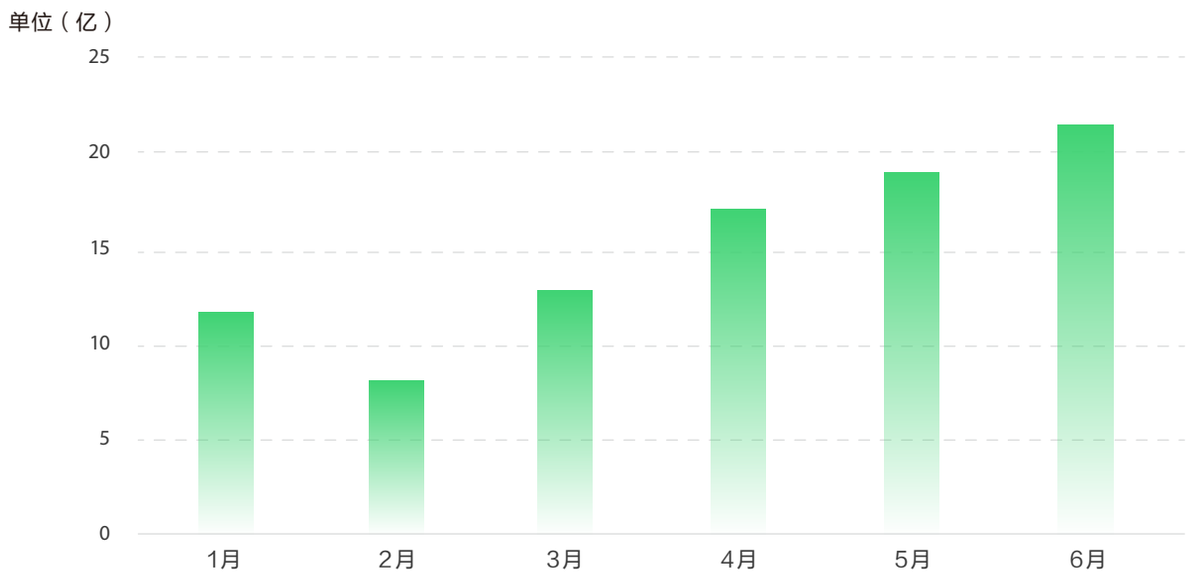


图2-4

同时，利用威胁情报能力把防护做到事前也是一个必然的趋势，特别是在云上，天然的大数据优势可以为威胁情报提供很好的基础。根据阿里云WAF全网协同防御数据所生成的威胁情报库看，每天在云上进行大量web攻击行为（从来不干好事）的IP超过了2000个，这些动态IP形成的威胁情报库每天拦截的攻击量超过了1600万条。

2. CC和爬虫攻击情况

从上半年云上的CC攻击情况看，在攻击事件数量上基本与去年保持持平，虽然最大的CC攻击事件攻击峰值超过了200万QPS（每秒请求数），但我们看到大流量暴力型的CC攻击在减少，并且攻击变得更加灵活和智能，在一些高级攻击中往往伴随有较为完善的监控体系，在被拦截后能够做一定程度的攻击手法变换以达到动态对抗、绕过防护策略的目的。而相应的，企业也应该在CC防护上具备自动对抗的能力，当前阿里云安全产品所使用的基于AI自动学习流量基线并生成防护规则的防护引擎就是我们在这一领域的最佳实践。

攻击者正在向更精细化的方向演化，更多类似CC攻击的行为出现在一些业务属性明显的接口上，如登录接口、短信验证码接口、查票接口、专利查询接口等。即越来越多的CC攻击是由于爬虫爬取量过大带来的“附加伤害”所引起的拒绝服务效果。

反观爬虫，时至今日爬虫已经有了明显的产业化，背后有明确的利益驱动，可能是营销作弊、倒卖低价票、爬取价值资讯、薅羊毛等。这类爬虫往往会对一些特定的业务感兴趣（如查询机票），而且往往会规模化、常年持续性的爬取这类业务接口以获取最新的价值资讯（如航线和票价）。在我们统计到的一些用户的票务接口上，来自爬虫的日常流量占比已经超过了98%，相应的正常用户请求只有不到2%。

同时，秒拨IP和代理IP已经被大规模的应用，基于我们对云上比较受爬虫“青睐”的业务观察，单个域名24小时内爬虫所使用的去重IP数量就达到了80万以上，而这个数字在2018年仅为27万。更加值得关注的是，第二天被复用的爬虫代理IP中跟前一天的重合度只有38%左右。

这两个数字意味着传统的固定IP池已经在向大规模的动态IP池进化，同时爬虫所使用的代理IP和正常用户所使用的IP重合度也越来越大，如果单纯的以IP作为客户端标记进行检测和拦截，将会带来越来越多的漏报和误伤。

另外，专业爬虫团伙所使用的代理IP池已由国内向全球各大洲转移，云上统计到的爬虫IP来源地中，中国大陆地区仍占了超过半数。除中国大陆地区外，在东南亚、北美和欧洲也有着大量分布，如图2-5所示。

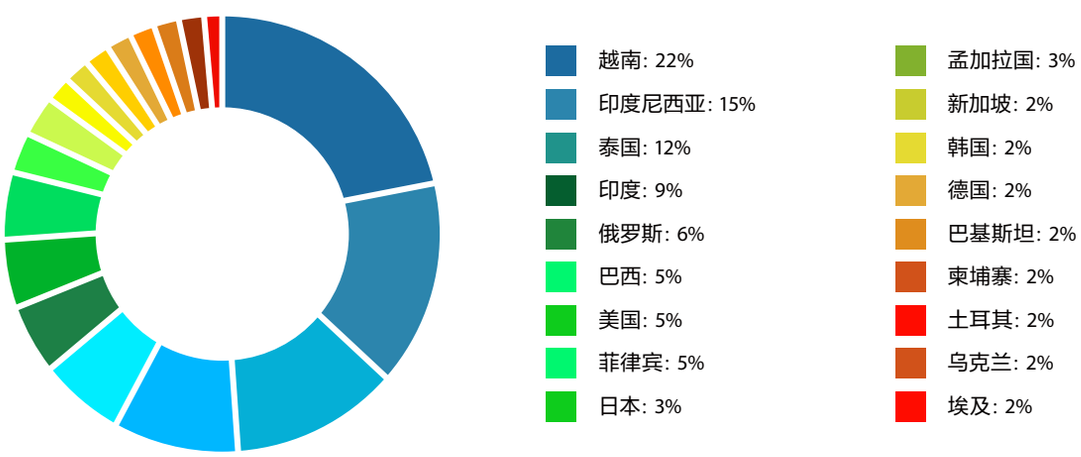


图2-5

03

核心观点与安全建议

VIEWS AND SUGGESTIONS

1. 90%以上攻击流量来源于扫描器

扫描器往往是攻击者的开路利器，在大规模批量扫描中被嗅探到大量漏洞的web站点更容易成为攻击者下手的对象。通过特征、行为等维度识别并拦截扫描器请求，可以有效降低网站被攻击者盯上的概率，同时有效缓解批量扫描行为带来的负载压力。

从目前的数据来看，拦截的攻击中，扫描器产生的请求数量在90%以上，除去扫描器自动化产生的攻击，剩下的10%手工测试行为，0day，广度低频等攻击则是需要花上90%精力来解决，如图3-1所示。

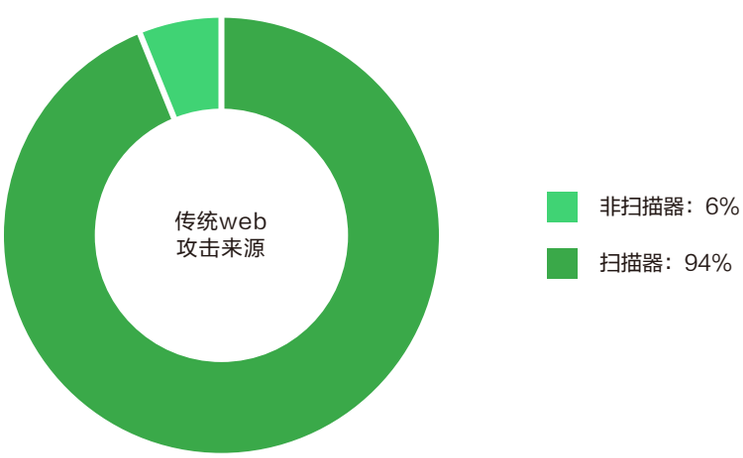


图3-1

2. 利用编码绕过防护的行为愈发普遍

随着WAF对网站的防护越来越普及，针对基础web攻防来说，利用诸如MySQL、JavaScript语言特性进行各种编码、变形，从而绕过WAF防护的攻击payload也越来越多，攻防是一个持续对抗升级的过程。根据云上数据显示，当前已有近1/3的攻击数据采用了不同程度或类型的编码、变形手段，以期绕过云盾WAF的防护，其中甚至不乏使用多维度的复合变形、编码手段实施攻击。

云盾WAF新一代引擎架构，支持多种常见HTTP协议数据提交格式全解析：HTTP任意头、Form表单、Multipart、JSON、XML；支持常见编码类型的解码：URL编码、JavaScript Unicode编码、HEX编码、Html实体编码、Java序列化编码、base64编码、UTF-7编码；支持预处理机制：空格压缩、注释删减，向上层多种检测引擎提供更为精细、准确的数据源。

该架构主要特征包括：在准确性上，优化引擎解析HTTP协议能力，支持复杂格式数据环境下的检测能力；抽象复杂格式数据中用户可控部分，降低上层检测逻辑的复杂度，避免过多检测数据导致的误报，降低多倍的误报率；在全面性上，支持多种形式数据编码的自适应解码，避免利用各种编码形式的绕过。

3. IP身份不再可信

IP地址是传统防护中一个非常重要的手段，很多经典的防护手段，如限速、名单、异常行为识别、威胁情报等都是基于IP地址实现的。但随着现在黑灰产对大规模代理IP池，特别是秒拨IP的广泛使用，IP地址已经变得不再可信。同一个IP地址，在10分钟前还被合法用户小白用于浏览A网站，在10分钟后已经被黑产人员小黑用作撞库攻击的代理IP，一个IP背后的身份开始变得极其复杂，黑与白交接的灰色地带比例在迅速扩大，这对于很多传统安全方案（不论是黑名单机制还是白名单机制）都带来了颠覆性的威胁，带来的相应误报和漏报也在迅速增长。

相应的，防护一方也应该做出改变。我们建议在做安全防护方案时，一方面将IP的身份或信誉辅助以其他维度的情报信息或者二次校验手段综合判断；另一方面降低对于IP的依赖，从更多维度去标识一个“客户端”或者“用户”，如设备指纹、业务中打点的token、cookie等等。

04

漏洞应急响应实践案例

PRACTICAL CASES

对于新爆发漏洞的应急响应和自动防御是云WAF的核心能力之一，上半年阿里云安全团队共进行了13次针对新爆发web漏洞的应急，这其中最具代表性的是weblogic远程代码执行漏洞。

Weblogic是Oracle公司出品的著名中间件产品，weblogic的XMLDecoder出现了多次远程代码执行漏洞，官方修复的方式一直是采用黑名单绕过的方式，已经成为了黑客的主要攻击目标，本次应急分析的Weblogic远程代码执行漏洞(CVE-2019-2725)，由于采用的是HTTP协议发送请求，所以从云上监控数据来看，黑客更喜欢利用通过HTTP协议执行的漏洞进行挖矿、勒索等目的攻击。而weblogic的主要使用用户多为保险、证券、网贷等互联网金融行业，这些客户的数据更加敏感和重要。

2019年4月17日，CNVD公布编号为CNVD-C-2019-48814(CVE-2019-2725)的WebLogic漏洞，指出该漏洞受影响的war包为bea_wls9_async_response.war。wls9-async组件为WebLogic Server提供异步通讯服务，默认应用于WebLogic部分版本。由于该war包在反序列化处理输入信息时存在缺陷，攻击者通过发送精心构造的恶意HTTP请求，即可获得目标服务器的权限，在未授权的情况下远程执行命令。

阿里云WAF团队在监测到该漏洞后立即进行分析，发现除bea_wls9_async_response.war之外，wls-wsat.war也受到该漏洞影响。

4月23日CNVD追加通告称，该漏洞受影响的war包不仅仅包括bea_wls9_async_response.war，还包括wls-wsat.war。该war包提供了WLS-WebServices的路由，而WLS-WebServices功能使用了XMLDecoder来解析XML数据。

4月21日，阿里云针对该漏洞更新了默认防御规则，开启拦截模式，实现用户域名接入即可防护。通过对阿里云上的流量监控发现在Weblogic远程代码执行(CVE-2019-2725)漏洞爆发后的攻击趋势如下图所示：

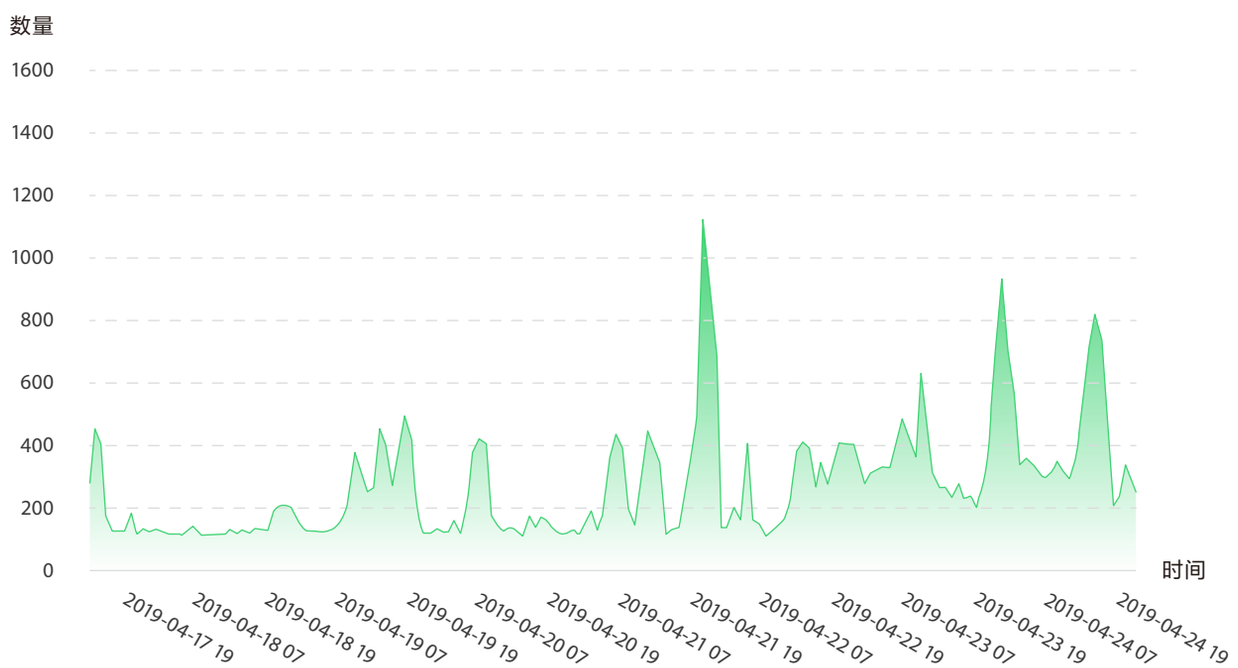


图4-1

从漏洞类型看，主要高危漏洞都是属于远程命令执行(Remote Code Execution)漏洞，其中不乏国内外主流的Web中间件和Web CMS（如Weblogic、Jenkins、Drupal、ThinkPHP），从防御规则更新后在云上实际的拦截量来看，ThinkPHP5.0.x远程代码执行漏洞以绝对优势占据榜首，其他规则的拦截量分布如下：

拦截量

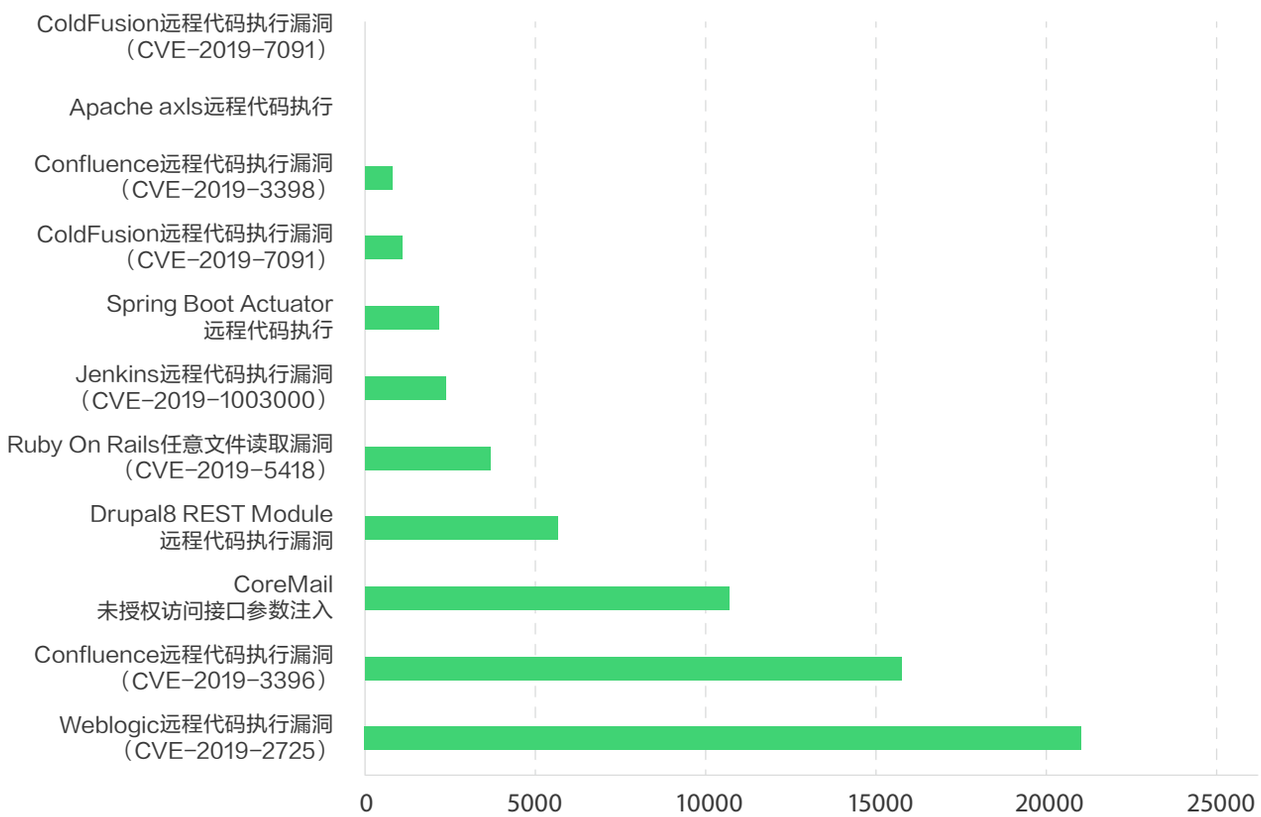


图4-2

以上就是阿里云针对上半年云上web安全情况的分析和解读，希望对各类站点的运维人员和安全人员带来参考，共建安全的互联网！

阿里云安全团队