腾讯安全
Tencent Security

揭秘BOT流量防范新型攻击

腾讯安全 BOT 管理白皮书

序言

BOT 流量是指在互联网上对 Web 网站、APP 应用、API 接口通过工具脚本、爬虫程序或模拟器等非人工手动操作访问的自动化程序流量，一般也称为机器人流量。据第三方调研报告统计，2021 年的 BOT 流量请求占比已经超过人工的访问流量。而 BOT 流量也与我们日常生活密不可分，不管是抢票抢菜，还是领券带货，我们甚至在不经意中就参与其中。当然，BOT 流量也并非都是恶意的，也存在良好 BOT 流量，如搜索引擎、统计和广告程序等正常流量能提升网站排名，进行网站监控提升用户体验。恶意的流量通过利用代理或秒拨 IP、手机群控等手段来爬取信息数据、抢刷接口、薅羊毛、外挂作弊等恶意攻击行为，对业务带来信息泄露、资金损失等风险损害网站和用户的利益。

《白皮书》将从流量构成和攻击特征帮助企业了解和认识 BOT 流量，并深入剖析常见的 BOT 类型、使用场景和恶意 BOT 的危害。同时，《白皮书》还将介绍业界主流的 BOT 攻击对抗方案，为企业提供恶意 BOT 流量防护思路。最后整体分析 BOT 的市场规模及发展趋势。

《白皮书》分为四个部分，第一部分，从流量构成和攻击特征介绍 BOT 流量；第二部分，分析常见的 BOT 类型、使用场景以及恶意 BOT 的危害；第三部分，全面介绍业界主流的 BOT 攻击对抗方案，为企业提供恶意 BOT 流量防护思路；第四部分，分析 BOT 未来的市场规模及发展趋势。

目录

2022 年上半年 BOT 流量分析主要观点

BOT 流量占比逐年上升 \ 02
BOT 攻击产业化、普及化、自动化 \ 03

BOT 常见类型与对抗手段

BOT 常见类型 \ 07
BOT 主要对抗手段 \ 10

2022 年上半年 BOT 流量现状分析

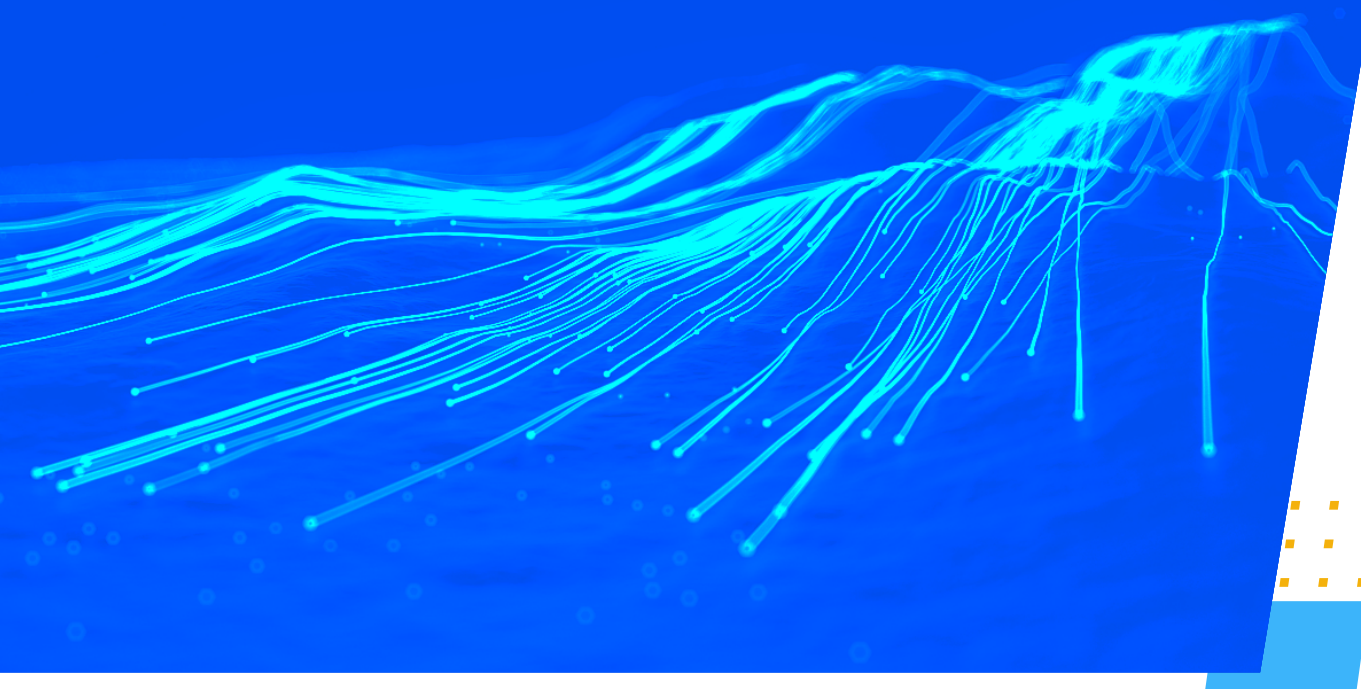
常规的 BOT 对抗方案

基于规则情报的 Anti-BOT 方案 \ 15
基于客户端风险的 Anti-BOT 方案 \ 16
基于机器学习+ AI 的 Anti-BOT 方案 \ 19
基于规则情报+客户端风险识别+机器学习+ AI 的 Anti-BOT 方案 \ 20

BOT 市场规模与行业分析

市场规模与预期 \ 22
疫情中 BOT 的趋势变化 \ 22
网络攻击成为 BOT 新兴攻击流量 \ 23
来自不同网络类型的流量分布更加均匀，来自基站的网络流量增加 \ 23
随着技术的不断迭代，滑动验证码在识别 BOT 流量上的效率有所降低 \ 24
游戏、零售和电子商务行业受到 BOT 攻击最多 \ 24

2022 年上半年 BOT流量分析主要观点



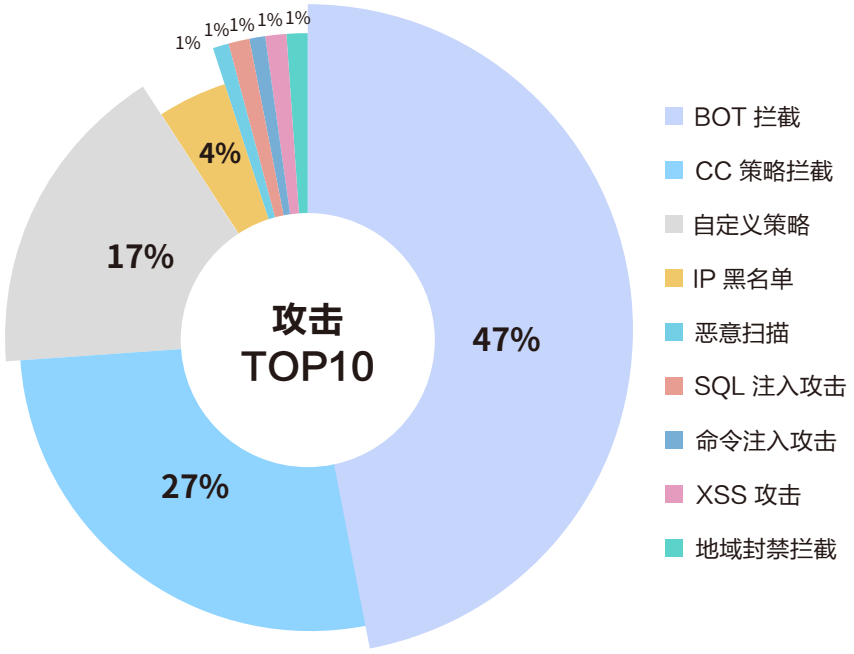
BOT流量占比逐年上升

2022年上半年平均每月BOT流量占整体流量63%，恶意BOT流量占整体流量27%，恶意BOT流量增长趋势迅猛多端混杂，攻击目标从业务资源型BOT逐步切换为针对业务内容的API型BOT，多端BOT流量混杂，对BOT防护的粒度有较大的要求。

- 2022年上半年平均每月的Web应用的攻击流量中，BOT与CC攻击流量占据整体网络攻击流量的80%，针对业务攻击流量远大于Web应用攻击流量，环比2021上半年的攻击流量数据，BOT攻击流量整体上涨幅度为5%。

  2022年上半年BOT攻击流量平均每月达到110亿次数+攻击流量，CC攻击流量为63亿次数攻击流量。现网上的主要攻击流量类型以业务攻击流量为主。

- BOT自动化攻击流量不再仅伪装浏览器发起，而是在多端混杂上更进一步。随着居家办公及移动办公的普及，Web应用上的流量不再仅仅局限于浏览器。小程序、APP逐渐成为新生的流量载体，BOT流量也随着时代开始改变，BOT自动化的攻击流量不再局限于伪装浏览器，网多端混杂更进一步。



BOT 攻击产业化、普及化、自动化

2022 年上半年 BOT 上下游产业链密切配合持续丰富，攻击者提供的攻击服务产业化，攻击者形成 BaaS（BOT as a Service）趋势。

BOT 攻击者的上下游供应链继续丰富，云上提供的相关服务内容增多、除了攻击者常用喜爱的 IDC、VPS 此类较传统的机器外，可选择使用路径更加多，如近几年新兴的云函数、Serverless、无服务计算、云真机等等技术发展的兴起，部分攻击者使用的资源/机器资源切换手段从老式的自己购买 VPS / IDC 搭建基础环境，变换为使用云函数、Serverless、无服务计算、云真机进行低成本的机器资源的模拟及使用，并形成相关的服务信息。

BOT 使用供应链相关资源配置的丰富外，很多攻击者会使用一些来自商业化的配置工具进行访问，如代理服务商、VPN 服务商、模拟器服务商、沙盒服务商等上下游资源供应链。除了自建业务应用外，使用上下游成熟的业务也不在少数，BOT 攻击者通过购买现有商业化的资源替换方案，通过商业化的模拟器、沙盒、IP 代理，实现业务资源的快速 Anti-BOT 对抗。

部分 BOT 攻击者整合了上下游供应链的信息，通过低代码的方式，为有需要的攻击者提供自动化 IP 变换、自动对抗验证码、自动化沙盒等对抗技术，分摊降低 Anti-BOT 对抗的成本，实现 BaaS（Bot as a Service 成为新宠儿）化服务方式。

2022 年上半年 BOT 攻击的使用手段及技术更加普及，BOT 流量的发起也不再局限于灰黑产业务中。

随着信息传播的加速，BOT 利用工具也在不断的传播，BOT 工具使用人员不再局限于灰黑产业务人员中。

疫情下，网络空间的流量发展更进一步，很多业务数据从线下实体，延伸至线上服务。从之前的火车票抢订、酒品抢购、医院挂号再到生活物资的购买，都逐渐从线下逐步切换到了线上，“黄牛党”、“羊毛党”等使用 BOT 技术的人员群体为主要 BOT 流量发起者，但是随着时间的逐渐迁移，部分“黄牛党”将 BOT 工具通过分销的形式进行售卖，并提供相关技术支持，部分正常用户也可以通过利用这种 BOT 程序，对业务进行恶意 BOT 访问。

信息技术不断发展，打造 BOT 工具的门槛持续降低，部分用户选择自建 BOT 工具发起 BOT 流量。

随着计算机信息技术的不断普及传播，部分恶意用户尝试通过已有的技术，自己编写 BOT 相关工具对业务进行重复性 BOT 访问，并将这类 BOT 工具在公开代码平台、社交平台上进行传播，使得部分正常用户也会使用 BOT 工具，在没意识到是攻击行为的情况下对业务进行 BOT 访问和攻击，从而影响了业务的正常运行。

2022 年上半年 BOT 技术手段变化多样，恶意 BOT 流量的识别和防护难度增加。

基础资源调度更加便利，随着 win11 WSA（Windows Subsystem Android）这个新特性的发布，更多的攻击者通过利用 WSA 对业务进行拟真访问，解除了之前的传统模拟器中系统应用版本低的问题，BOT 流量识别难度加大。

技术手段更进一步，随着 Chrome Devel-op Protocol、Playwright 等操控自动化操作工具的逐渐完善，在端上识别浏览器被操控的难度逐渐增加，如果要完全处置此类 BOT 手法，会造成在移动端上的大量碰撞，误伤正常用户，BOT 识别难度加大。

最近一年中 BOT 攻击者的发起网络请求的位置更加偏向从住宅 IP 中发起，IP 类型从传统的 IDC、VPN 侧发起，逐渐变为混淆度极高的住宅 IP、基站 IP。BOT 攻击流量混淆在里面,处置攻击者的同时更容易误伤正常访问用户。

BOT 相关工具广泛传播，正常用户也会使用 BOT 相关工具，如果配置不当，易造成相关客情舆论，造成防护规则难度加大。

2022 年上半年 网络攻击类更加自动化、武器化。

网络攻击者利用 BOT 对网络空间进行大面积的扫描攻击，从漏洞 POC 发现到 BOT 的自动化利用时间间隔大幅度降低。需要强有力的对抗自动化网络武器的手段。

漏洞爆发迅速，攻击者利用 BOT 工具对网络进行大规模扫描，在漏洞爆出的初期，快速实现 Web 应用攻击扫描，如在 2021 年底爆发的 Log4j2 漏洞，攻击者在漏洞公布后的几个小时内就已经开始全网大规模的扫描。除了基础安全规则的防护外，需要一个更加有力的手段，用于对抗此类自动化的批量扫描工具。

除此之外，攻击者为了达成目的，会使用自动化模糊测试的手段，对敏感业务的接口字段进行安全防护绕过，以获取相关的敏感业务资产信息。如果对此类绕过探测手段没有比较好的防护措施，将会使敏感的基础设施的权限、数据敏感信息被黑客窃取，造成业务资产损失。

BOT 常见类型与对抗手段

BOT常见类型

爬虫机器人

爬虫机器人，也称为网络蜘蛛或爬虫，通过跟踪超链接浏览网络，目的是检索和索引网络内容。蜘蛛下载 HTML 和其他资源，例如 CSS、JavaScript 和图像，并使用它们来处理站点内容。

如果您有大量网页，您可以将 robots.txt 文件放在您的网络服务器的根目录中，并通过自定义设置并向爬虫机器人提供说明，指定它们可以抓取您网站的哪些部分以及频率。

抓取机器人

抓取机器人是从网站读取数据的机器人，目的是离线保存数据并使其能够重复使用。抓取机器人可能抓去网页的全部内容或特定的 API 数据以获取特定的数据，例如电子商务网站上产品的名称和价格以及详情图片。

网页抓取是一个灰色地带，在某些情况下抓取是合法的，并且可能会得到网站所有者的许可。在其他情况下，机器人操作员可能会违反网站使用条款，或者更糟糕的是利用抓取来窃取敏感或受版权保护的内容。

垃圾邮件机器人

垃圾邮件机器人是一种互联网应用程序，旨在收集垃圾邮件列表的电子邮件地址。垃圾邮件机器人可以利用电子邮件地址的独特格式从网站、社交媒体网站、企业和组织收集电子邮件。

在攻击者积累了大量电子邮件地址/或使用临时邮箱后，他们不仅可以使用它们发送垃圾邮件，还可以用于其他邪恶目的：

凭据破解	将电子邮件与常用密码配对，以获取未经授权的帐户访问权限。
表单垃圾邮件	自动将垃圾邮件（例如广告或恶意软件链接）插入热门网站的表单中，通常是评论或反馈表单。

社交媒体机器人

社交媒体机器人在社交媒体网络上运行，用于自动生成消息、倡导想法、充当用户的追随者，以及作为虚假账户自己获得追随者。社交机器人可用于渗透人群并用于传播特定想法。由于对其活动没有严格的规定，社交机器人在网络舆论中扮演着重要角色。

社交机器人可以创建虚假帐户（尽管随着社交网络变得越来越复杂，这变得越来越困难），放大机器人操作员的信息，并产生虚假的追随者/喜欢。很难识别和缓解社交机器人，因为它们可以表现出与真实用户非常相似的行为。

控评	通过利用大量注册的虚假账户，对社交媒体中的相关评论区进行刷屏控评，控制相关舆论信息。
定向引流	通过利用大量注册的虚假账户，对社交媒体中的相关评论区进行特定数据引流，引导正常用户到宜昌的页面中。

下载机器人

下载机器人是可用于自动下载软件或移动应用程序的自动化程序。它们可用于影响下载统计数据，例如在热门应用商店获得更多下载，并帮助新应用登上排行榜榜首。

它们还可用于攻击下载站点，创建虚假下载作为应用层拒绝服务（DoS）攻击的一部分。下载机器人通过创建下载链接，影响业务带宽，造成正常用户的不可访问，影响正常业务访问。

狙击手机器人

狙击手机器人是一种自动购买热门活动门票、购买热销商品、热销旅游票务的方式，目的是转售这些商品以获取利润。这种活动在许多国家都是非法的，即使没有被法律禁止，对活动组织者、售票者和消费者来说也是一种烦恼。

Sniper Bots 往往非常复杂，会模仿人类抢购的行为。在许多抢购领域，自动机器人购买的票的比例在 40–95% 之间。

攻击机器人

攻击机器人用于在漏洞爆发时期，被攻击者载入攻击载荷从而实现大规模的 Web 应用漏洞攻击的机器人，用于攻击互联网上所有 Web 应用的站点，已获取其站点的系统权限、数据资料信息。最终用于肉鸡、傀儡机、勒索等威胁网络、系统安全的机器人。

账号接管机器人

账号接管机器人是用于不断的进行业务账号爆破的机器人，主要用于不断爆破当前站点的应用账号信息,用于获取当前站点的所有账号资料，类比行为：撞库、爆破，账号接管机器人最终得出成果大多数用于灰黑产的账号交易。

扫描机器人

扫描机器人是互联网中无差别扫描 web 站点的机器人，通常这些机器人用于收集网络空间资产信息，常用于灰黑产、攻击方的信息收集。

BOT 主要对抗手段

BOT 技术在过去数十年间不断发展变化，其目的以及技术手段都也不断的发生改变。在最初的阶段，BOT 技术仅用于检索数据或执行操作，其本质为脚本工具，这些脚本不接受 Cookies 也不能解析 JavaScript 。因此脚本特征较为明显。较容易进行检测及对抗。

随着时间的推移，BOT 所使用的技术及目的也变得越来越复杂，出现部分 BOT 使用的技术不仅仅可以接受并存储使用 Cookies 的技术，还可以动态解析网站下发的 JavaScript 脚本、CSS 动画渲染的内容，以加载需要的动态渲染的网站内容，获取更多可以获取的业务数据内容。但是这种可以主动解析JavaScript 的仿真浏览器类型的爬虫，仍然可以比较快速的发现及对抗。在业务环境中，正常用户使用浏览器和仿真浏览器对页面内容进行访问解析，在可视元素以及渲染加载元素中会存在相关页面差异。可以通过类似敲门的功能，检测是否为仿真浏览器进行访问。此外，也有部分攻击者通过解析 JavaScript 内容，利用脚本工具仿真加密协议，对网站业务进行访问。

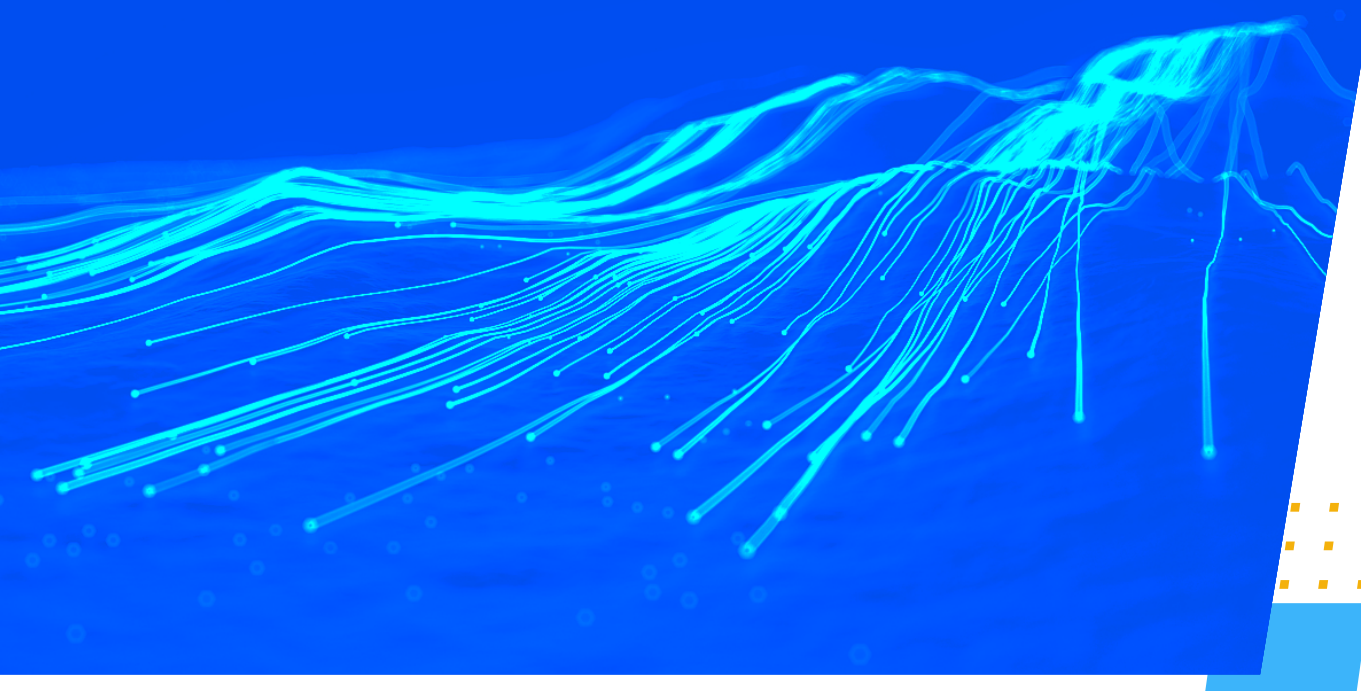
近两年使用像 PhantomJS / Headless 这样的无头浏览器——这些浏览器可以完整地处理网站内容。与真实用户几乎没有区别。这些机器人甚至可以模拟人类活动，例如点击页面元素。但由于存在端上的细微特征差异，可以被客户端风险识别识别出来，同时因为存在机械重复动作，会被大数据后端分析，精准的识别出来为异常用户。

随着模拟器、云真机、群控等灰黑产使用的商业工具链的完善，近年来 BOT 的拟真、对抗手段越来越多。同时随着近年移动互联网的高速发展，基于移动端上的 BOT 流量越发强烈，传统的 BOT 对抗不在仅仅局限于浏览器。现在攻击者更多偏向于使用移动端设备进行攻击。通过模拟真实设备，进行对抗。

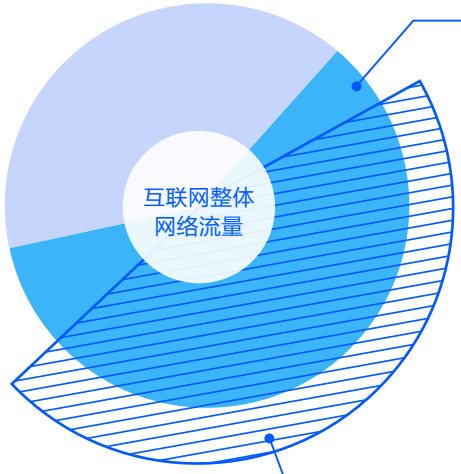
BOT 对抗的上下游资源链

技术准备	IP代理池	真机模拟	真机驱动	众包工具
资源准备	IP资源：代理 / 秒拨 / VPN / IDC / VPS / SCF	使用资源：浏览器 / 脚本工具 / 协议模拟 / 云真机	账号资源：账号注册 / 手机小号 / 身份证 / 邮箱小号 / 银行卡	
对抗准备	IP代理池	接码平台	解码平台	验证码对抗
	平台攻击	接口攻击	订单攻击	

2022年上半年BOT流量现状分析



BOT流量态势〔黑白灰〕

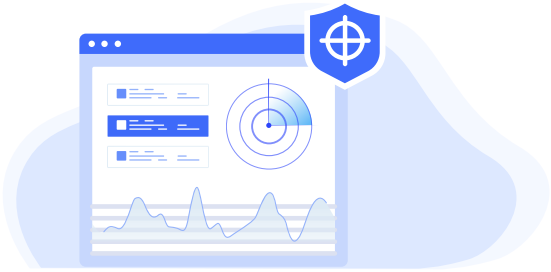


总体 BOT 流量
占整体互联网流量
约 60%

具备恶意攻击性的 BOT 流量占
据互联网整体网络流量中的
46%

网络攻击者通过 BOT 手段将攻击自动化、武器化

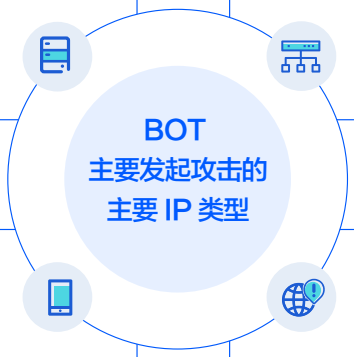
网络攻击者研发了许多自动化的网络攻击扫描工具，在漏洞爆发前期通过将 payload 放入自动化扫描工具中进行大量分布式扫描攻击



网络攻击者在进行定向攻击之，运用到了大量的自动化模糊测试工具对单一 API 进行攻击

BOT 主要发起攻击的主要IP类型

IP 地址是网络接入的载体，是有限资源，且每个 IP 都有较容易获取的公共属性数据，不易被伪造，因此黑灰产通过 VPN、代理、VPS、ADSL 混拨等方式隐藏真实访问 IP，通过不停切换 IP 出口制造全国用户访问的假象。



IDC / VPS 混拨

远程在 IDC/VPS 云服务器上架设多根网线，软件实现多线混拨，远程控制 VPS 服务器拨号，IDC/VPS 混播 IP 池资源丰富，可以实现跨城市切换,而旦速度快、稳定性好，是目前黑灰产团伙使用的主流 IP 切换模式

VPN / 网络代理

攻击者通过代理 IP/VPN 访问其他网站，可以掩盖流量的真实来源。需要 BOT 使用团队进行相关代理池 / VPN 池进行维护

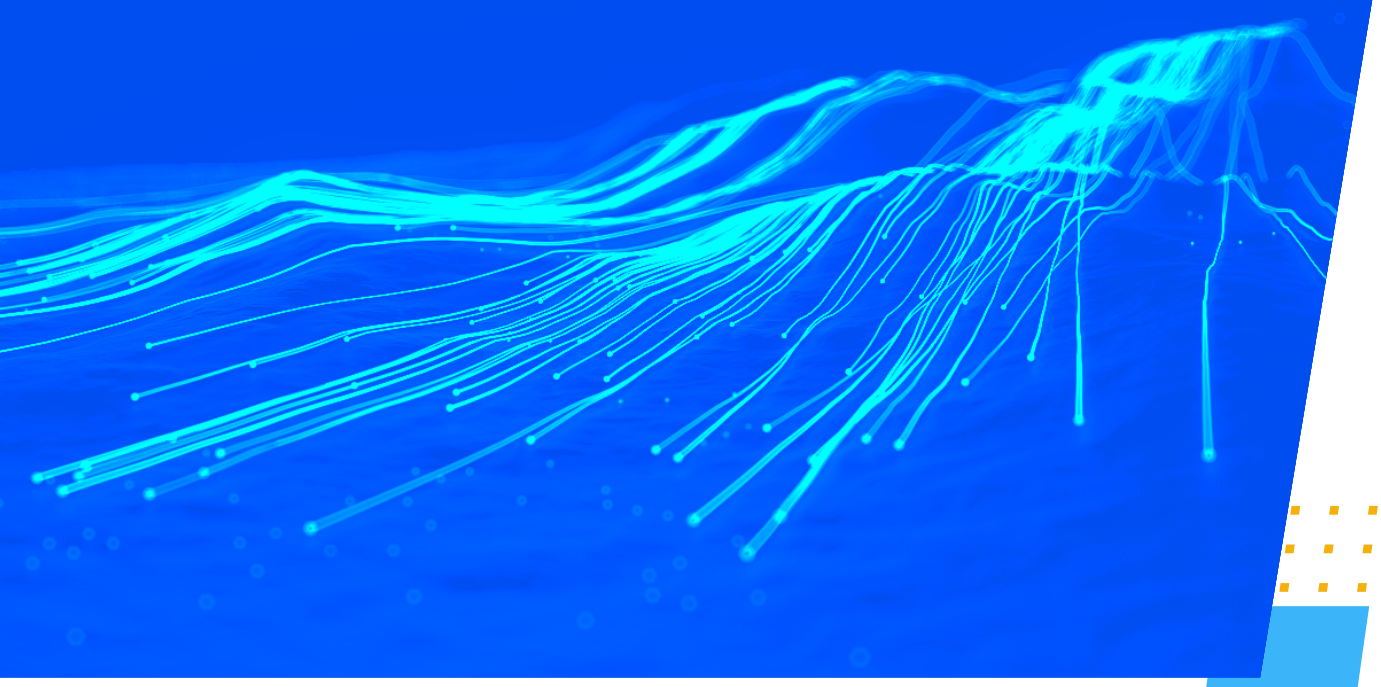
物联网基站 / ISP 住宅代理出口

攻击者通过物联网基站 / ISP / 住宅代理出口访问业务网站，通过将攻击流量混杂在正常流量中的。使进行业务处置的时候难以判断

秒拨

秒拨的底层思路就是利用国内家用宽带拨号上网（PPPoE）的原理，每一次断线重连就会获取一个新的 IP。与时俱进的黑产掌握大量宽带线路资源，利用虚拟化和云计算的技术整体打包成了云服务，并利用软路由对虚拟主机以及宽带资源做统一调配和管理

常规的 BOT 对抗方案



基于规则情报的 Anti-BOT 方案

Robots.txt

Robots.txt 是一个古老的爬虫协议文件，他的位置位于域名根目录下。譬如 http://example.com/robots.txt。严格来讲 Robots.txt 并不算一个反爬虫技术，而是一个由爬虫遵守的协议。它通过几个简单的命令告知遵守 Robots.txt 的爬虫哪些可以被爬取，哪些不能。

在内容的具体构成上，“爬虫协议”通常由一个或多个语法单元组成，每个语法单元可进一步分为 两部分：一个是 User-agent 值，用于设置其所允许或禁止的搜索引擎，后接其所针对的爬虫程序的名称；另一个是 Allow 或 Disallow 值，用于设置特定爬虫程序所能访问或禁止访问的具体路径。

一般的搜索引擎爬虫会遵守这个协议，而对于爬虫技术对抗的层次来说，这个文件毫无意义。

IP 层/网络层

IP 报文带有的最重要的信息就是 IP 请求的来源地址， 来源地址极难伪造的特性，使得这个字段成为 Anti-Bots 策略中最重要的字段。封杀 IP / IP 段是网站可以执行的最严厉的惩罚。由于国内的 ISP 大量的使用了 NAT 技术，多用户共用同一 IP 的情况越发常见，如果单独对 IP 进行处置，易产生误杀，影响正常用户的网站访问。但是即使如此， 源 IP 也是 Anti-Bots 策略中最为核心的数据，常规的 Anti-Bots 策略的处置一般都要围绕源IP进行，如针对异常访问 ip 、代理 ip、idcip 等。

HTTP 协议层有几个有趣的 HTTP 头，它们是制定反爬虫策略的常用数据。

HTTP 头部信息

Referer

Referer 是浏览器在页面跳转时带入的 HTTP头 部信息，用于表示 用户是从那个页面上访问进来的，可以根据 Referer 信息来定位用户访问的网页位置。 一般来说，PC 端 网站 90% 以上的 Web 请求流量应包含 Referer 字段。 在一些常见的 Anti-Bots 策略 中， 大量的不带 Referer 请求头、非本站来源 Referer 的访问请求会触发验证码策略。由于 HTTP 协议的特性，许多攻击者会模拟并伪造 Referer 来源于本站的请求，用于绕过简单的 Anti-Bots 策略。这也就是典型的 Referer 滥用场景。

X-Forwarded-For

HTTP 头部信息中 X-Forwarded-For (XFF) 字段是在客户端访问服务器的过程中如果需要经过 HTTP 代理或者负载均衡服务器，可以被服务器用来获取最初发起请求的客户端的 IP 地址。

XFF 会被用来进行调试和统计，以及生成基于位置的定制化内容，按照设计的目的，它会暴露一定的隐私和敏感信息，比如客户端的 IP 地址。 由于 HTTP 协议的特性，攻击者可伪造 XFF 字段信息，对网站进行伪造访问，绕过传统 Anti-Bots 的对 IP 的封堵处置策略。

User-Agent

User-Agent 首部包含了一个特征字符串，用来让网络协议的对端来识别发起请求的用户代理软件的应用类型、操作系统、软件开发商以及版本号。

知名 Bots 均有自己唯一的 User-Agent 信息，如 搜索引擎爬虫。搜索引擎爬虫通过 在 User-Agent 字段中标记自身所属的身份信息，告知 Web 应用服务器中的身份信息，方便 Web 应用提供相关的快速索引信息。但是由于 HTTP 协议的特性，有不少攻击者会尝试伪造 User-Agent 字段信息，用于冒充搜索引擎爬虫，对网站的业务进行访问爬取，这也是典型的 User-Agent 伪造的场景。

基于客户端风险的 Anti-BOT 方案

JS 渲染（Ajax 与 SPA）

Ajax 应用可以仅向服务器发送并取回必须的数据，并在客户端采用 JavaScript 处理来自伺服器的回应。由此，single-page application（SPA）页面应用也逐步开始盛行，许多页面内容通过 Ajax 进行动态获取与渲染。脚本类型的 BOT 若不未经定制化改造，是较难获取到对应动态页面加载渲染的数据，并且如果页面接口进行变化，定制化的脚本工具也需要同样进行变换。

接口加密与 JS 混淆

Ajax 接口 默认返回的是规整化的接口，返回的数据格式如 JSON/XML 数据。对于正常用户较为难读，但是对 Bots 或仅收集特定数据的带来了更多的便利。拥有解析前端能力的 Bots 工程师可以通过只需一点点的前端逆向能力，利用开发者工具，分析网络请求，就可以找到相关的 API 接口，即可通过对应的库解析出数据。

但是如果前端通过 使用 JavaScript 、WASM 等技术进行传输数据加密混淆、并把通过相关加密方法（如 DES、AES、RSA 等变换）进行数据传输接口进行加密的话，Bots 工程师对逆向难度将会增加。如果再增加 JavaScript 的相关 Feature 以及 Uglify 混淆压缩使 JavaScript 代码不可读，并加上相关 JavaScript 、WASM 代码加密，令 Bots 工程师无法轻松的逆向出加密计算的流程， 就可以达到一定的反爬目的。

但是在客户端侧，为了能正常实现业务逻辑，仍然需要展示出来，因此就有攻击者利用 selenium、headless 的形式，模拟浏览器进行访问，通过渲染页面并运行相关 JavaScript 代码。从而绕过此类 Bot 对抗策略。

验证码

验证码 (CAPTCHA) 是一种古老而有效检测是否人类的一种方式。从最初的简单图形验证码，如数字验证码、字母验证码、到后来的中文验证码。到现代的行为验证码，短信验证码、VTT 验证码。验证码是应用层最普遍的人机对抗技术。对于一些简单的数字、字母验证码，行为验证码，随着近几年图像识别，机器学习、神经网络技术的高速发展，有技术人员通过上述技术训练出的对抗验证码的模型,其成功率可以达到 80~90% 。

因此更高阶的验证码也随之诞生，形如 VTT 验证码，行为动作验证码等。因此，也有灰黑产专门用使用人工打码平台来对接处理复杂验证码问题，所以单凭验证码很难有效处理 Bots 问题， 并且过多的验证码也会导致正常用户的体验受到影响。

内容混淆与假数据

Bots 主要动作为自动化的获取目标数据，但是有部分目标数据的可呈现内容为人观看，如文本、数字内容。如果直接在页面上进行展示，Bots 可解析对应的 DOM 节点数据，即可获取到相关内容。因此就有部分传统的页面策略采取了字体混淆、页面混淆的方式，对页面内容进行编码转换，后续通过加载对应字体的形式，加载成为人类可以正常阅读浏览的页面。用于对抗普通的脚本 bots 。但随着 OCR 的技术不断推进，此种对抗能力越来越弱。此种数据混淆减缓了页面加载的时间，也影响了正常的业务体验。

访问行为分析

访问行为分析的 Bot 对抗策略，可以分为两个方向：1. 用户请求访问行为；2.用户操作轨迹行为。这两者主要差异在于，访问请求的时间序与在页面浏览的操作序的差异。

如在进行页面下单的时候，正常用户访问请求会先到商品详情页中，在商品详情页添加至购物车或者直接唤起购买页面。但是在 Bots 中，这种访问行为的展示就变得有趣了起来，物品抢购下单的时候，Bots 将会直接调用购买下单接口，减少页面加载等相关耗时操作。

此外，像是在移动端上，默认人类进行点击的操作为 tap，但是 Bots 如果采取直接对 dom 进行处置，默认处置动作为 click。通过这种差异化信息，结合其他的 Anti-Bots 手段，就可以对 Bots 造成有效对抗。低级的行为分析基于规则，高级的行为分析基于 AI 评估与智能统计。

Cookies 与 Storage

HTTP Cookie（也叫 Web Cookie 或浏览器 Cookie）是服务器发送到用户浏览器并保存在本地的一小块数据，它会在浏览器下次向同一服务器再发起请求时被携带并发送到服务器上。通常，它用于告知服务端两个请求是否来自同一浏览器，如保持用户的登录状态。Cookie 使基于无状态的 HTTP 协议记录稳定的状态信息成为了可能。

Storage 分为两种类型：1.LocalStorage；2.SessionStorage。 LocalStorage 属性允许你访问同源的对象 Storage；存储的数据将保存在浏览器会话中。LocalStorage 类似 Session-Storage，但其区别在于：存储在 localStorage 的数据可以长期保留；而当页面会话结束——也就是说，当页面被关闭时，存储在 sessionStorage 的数据会被清除 。

通过 Cookie 和 Storage ，就可以跟踪用户的行为轨迹。但是 LocalStorage 由于为浏览器特性，因此，对对抗常规的脚本型爬虫，有较好的处置能力。

Navigator

Navigator 接口可以用来作为用户访问 User-Agents 的状态和标识。Navigator 允许脚本查询它和注册自己进行一些活动，以及操作系统，浏览器信息。部分 Anti-Bots 策略可以使用只读的 window.navigator 属性检索 navigator 对象，以获取操作系统、当前浏览器相关信息。

JavaScript 引擎指纹

不同的浏览器底层引擎在执行相同的 JavaScripts 代码的时候，会产生不同的代码结果。这样就可以通过下发并执行特定的代码片段（如 eval.toString().length、errFirefox 等），即可判断出当前浏览器是否为进行浏览器伪造。

Canvas 指纹

Canvas 提供了一个通过 JavaScript 和 HTML的 < canvas > 元素来绘制图形的方式。它可以用于动画、游戏画面、数据可视化、图片编辑以及实时视频处理等方面。Canvas 不仅局限于图片处理，它还能监听用户的键盘输入、鼠标移动、以及触摸事件。不同浏览器、操作系统、以及操作系统环境，会使得 Canvas 的同一绘图操作流程产生不同的结果。Canvas 指纹被所有主流浏览器支持，并且可以被大部分的 PC、平板、智能手机访问。如果是相同的运行环境，同一套 Canvas 操作流程会产生相同的结果。 浏览器指纹的优势是不需要浏览器保持本地状态，即可跟踪浏览器。这样当攻击者同时唤起多个受控浏览器时，可以快速发现请求均出现于同一机器。

系统指纹

系统指纹常用于识别当前访问客户端的相关系统信息，如水平陀螺仪、USB 接口信息等，在移动端上，可以检测当前访问的借口是否包含 水平陀螺仪的借口，用于检测是否为模拟器使用。此外在现代浏览器中，也有相关 USB WEB API 用于检测当前客户端是否有插入 USB 端口。通过这种处置策略，可以快速的校验出来当前客户端是否在 IDC、模拟器上。

SSL 指纹信息

通过提取 SSL 握手中的相关特征，利用 SSL 进行指纹识别。在使用系统默认特征库的情况下，SSL 指纹可以帮助识别操作系统。通过 SSL 指纹信息，我们可以快速识别出当前客户端的访问请求的是否伪造。

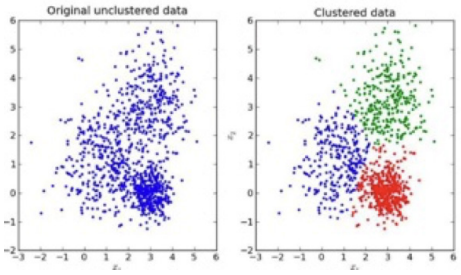
假链陷阱

假链陷阱通常为通过构造不可见的隐藏链接或不可被用户主动触发的事件嵌入在当前访问的页面中。由于大多数 Bots 的策略默认会解析当前页面的所有事件及链接，因此可以快速的发现此类 Bots。

基于机器学习 + AI 的 Anti-BOT 方案

会话访问行为特征

通过机器学习 + AI 防护的方式，计算出当前访问会话的相关访问行为特征，根据会话特征中的相关信息，如 URL 重复比、URL 种类、URL 平均深度、Cookie 是否滥用、Cookie 重复性、Cookie 有效率、User-Agent 类型、User-Agent 随机性指数、User-Agent 有效比、出现最多的 User-Agent 占比、Referer 重复比、Referer 存在比、Referer 有效比、出现最多的 Referer、出现最多的 Referer 的比例、请求参数比、请求参数种类对不同会话的访问行为进行处置。



会话访问意图特征

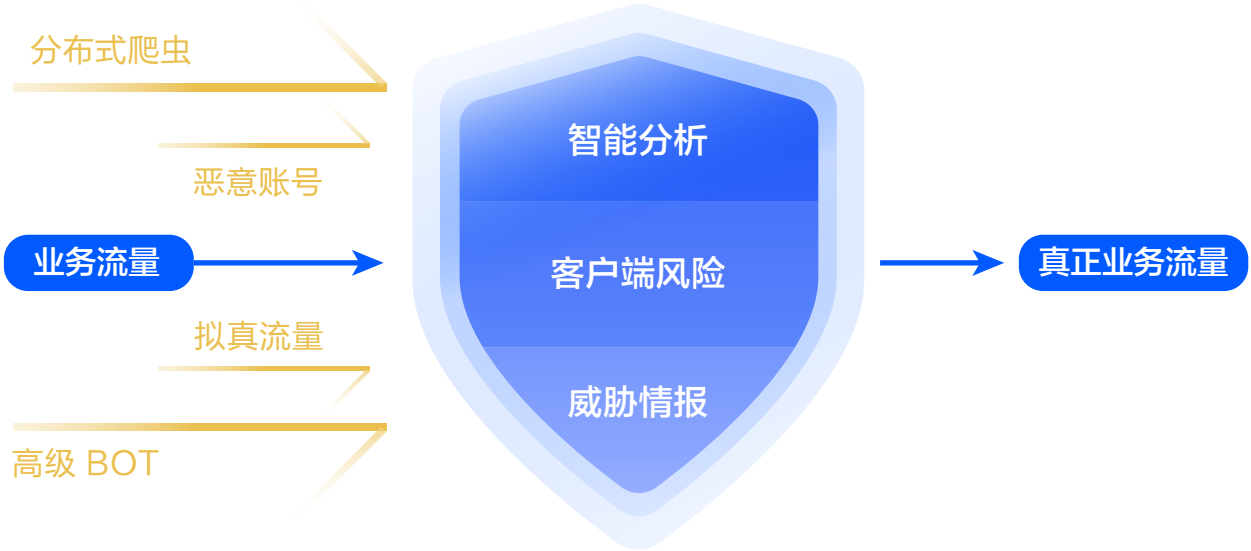
通过机器学习 + AI 识别的方式，计算当前访问会话的具体访问意图，并根据相关访问意图进行聚类，并形成会话访问行为意图聚集，并根据不同会话的意图规划进行聚类处置。

会话异常指标特征

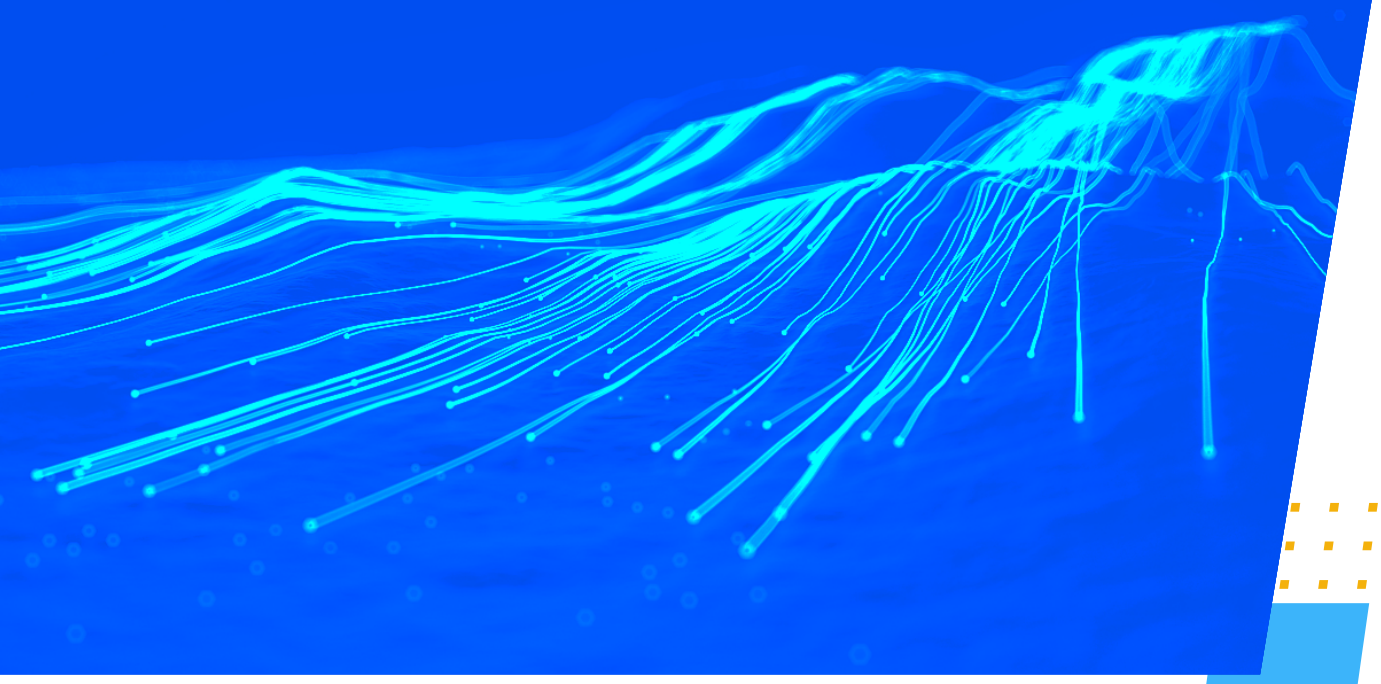
通过大数据统计可快速筛选出远超中位数的异常会话访问行为，通过相关异常访问行为指标，即可快速筛选出行为异常的流量。

基于规则情报+客户端风险识别+机器学习+ AI 的 Anti-BOT 方案

通过规则情报将存在异常的 IP（代理、扫描器、威胁情报）、BOT 访问特征进行快速过滤，随后通过客户端风险识别中的检测是否真人真机、最后通过后端的机器学习 + AI 方案分析得出异常的访问行为，并进行处置。



BOT 市场规模与行业分析



市场规模与预期

全球 Bot Management 的市场规模预计将从 2021 年的 4.08 亿美元增长到 2026 年的 9.83 亿美元，预测期内的复合年增长率 (CAGR) 为 19.2%。恶意 BOT 程序安全市场的主要驱动因素包括不良僵尸程序流量的增加；BOT 攻击的复杂性和组织的收入损失增加；BOT 浏览移动端访问量增大；电子商务、旅游、游戏等业务中 BOT 攻击激增。

疫情中 BOT 的趋势变化

新冠疫情影响了社会的各个行业，几乎所有个人和企业的生活方式都有一定的影响，互联网生态系统在全球范围扮演的角色越发重要。由于新冠疫情，人们对在线业务的依赖显着增加，导致恶意机器人流量的比例不断增加。在新冠疫情爆发后，2021 年 63% 的互联网流量访问不是人类由人类发起；恶意 BOT 流量增加了 4%，占所有网站请求的四 分之一以上。供应商对 BOT 管理的解决方案的需求也在不断增加。在新冠的影响下，拥有远程办公，数字化经营方式的能力已成为各种组织的主流要求。

随着移动设备和互联网应用在全球范围内的普及，企业逐渐倾向于使用 BOT 管理方案来保护免受 DDoS 攻击、数据抓取、账号爆破、垃圾邮件和其他恶意软件威胁。单需要注意的是，新冠疫情对 BOT 管理的市场产生了负面影响较大，由于大部分企业预算不足，BOT 管理软件的采购额均有一定程度的下降。

网络攻击成为 BOT 新兴攻击流量

网络犯罪正朝着利润驱动的方向发展，攻击者通过使用 BOT 对目标行业进行快速侵害网络犯罪分子、恶意软件运营商、工具提供商通过自己编写或利用相关的软件工具包，实现垃圾邮件发送、数据盗窃和执行 DDoS 攻击，从而可以轻松地从众多网站在线订购相关稀有资源商品。在世界各地观察到，BOT 针对个人或企业开展网络战活动的攻击趋势越发猖獗，发现部分企业难以应对这些 BOT 攻击，导致企业业务停滞。

来自不同网络类型的流量分布更加均匀
来自基站的网络流量增加

根据 IDC 数据，2021 全球智能手机总出货量为 13.5 亿部，发达经济体的智能手机拥有率明显更高。在线购物、社交媒体应用程 序和产品研究等各种活动都见证了智能手机的使用。QQ、微信、小红书、微博等社交媒体 应用程序以及腾讯视频、爱奇艺、优酷 等数字娱乐 平台的出现也使智能手机用户在他们的设备上花费更多时间。因此，网络流量从网络转移到移动设备， 使其成为 BOT 攻击的有利发起点。移动网络流量约占全球网络流量的 32.3%。

随着技术的不断迭代，滑动验证码在识别 BOT 流量上的效率有所降低

CAPTCHA 是一种非常重要的人机对抗方式，CAPTCHA 可以保护网站免受 BOT 和自动黑客工具的侵害。CAPTCHA 包括隐藏在需要手动验证的图像中的相关内容。尽管恶意机器人擅长自动完成表格，但对隐藏在图像中的语意进行识别对它们来说是困难的，因为人类可以阅读隐藏在图片中的语意，而计算机较为困难。但随着技术的进步，计算机变得越来越智能。黑客和灰产正在利用 AI 和 ML 技术，以便 Bots 可以自学如何分析图像并识别隐藏的意图。他们甚至可以准确识别图像中的特征元素，从而绕过较新的滑块 CAPTCHA 方式。因此，在当下环境 CAPTCHA 系统容易被这些训练有素的 Bots 绕过， 从而导致 BOT 攻击成功。

但是 CAPTCHA 的验证方式也在不断的提高，如 VTT 验证码或动态下发/经过混淆的滑块验证码的也在不断也起到了不错的对抗效果。

游戏、零售和电子商务行业受到 BOT 攻击最多

在线零售和电子商务业务主要是 恶意 Bots 攻击的目标，恶意 Bots 会执行 例如锁下单、抢购、数据抓取、订单填充和 7 层 DDOS 攻击。这些攻击可能导致网站访问不畅、网站停机、敏感客户数据泄露以及收入损失。同时，因为电子商务平台的可用性和安全性对于建立客户信任至关重要，因此，零售和电子商务的机器人安全解决方案预计在未来几年将具有 巨大的潜力。

腾讯安全
Tencent Security

业务咨询请联系

产品试用请扫码