CLOUDFLARE

白皮书

恶意机器人应对手册：预警信号以及处理方法

目录

引言

如今，机器人流量约占在线流量的 30%——其中许多机器人旨在对您这样的组织进行破坏——而且大约 93% 的机器人流量是未经验证且可能具有恶意的。内容和价格抓取、帐户接管、凭据和信用卡填充、库存囤积以及僵尸网络驱动的分布式拒绝服务 (DDoS) 攻击的日益普遍，表明恶意机器人行为者正逐年变得越来越复杂和老练。

此外，传统的反机器人措施（例如位置阻止、IP 地址阻止和传统 CAPTCHA 验证码）已不再有效。实际上，与人类相比，机器人更容易解决 CAPTCHA[1]。

因此，没有一种策略能够阻止每一种机器人，并防止其对您的用户和品牌造成伤害。唯一有效的方法是保持警惕，关注各种表明机器人问题的警示信号，然后做出响应：收集数据并部署有针对性的措施、模式检测、预测分析以及其他补充策略。

1. https://www.usenix.org/conference/usenixsecurity23/presentation/searles

存在机器人问题的警示信号

通过跟踪一系列潜在的指标，您很有可能在恶意机器人造成严重破坏前发现它们。需要注意如下迹象：

基础设施成本上升，但业务量并无增长

您网站的所有流量都需要一定的成本。无论是谁访问什么内容，您都必须承担存储和计算的费用。但是恶意机器人可能会增加与流量相关的成本，而不会给您的业务带来任何收入。虽然搜索引擎使用善意机器人来索引您网站上的内容，从而支持您的 SEO 排名，但恶意机器人每年都会造成高额带宽费用。

针对量少、高需求库存的异常购买

如果您发现库存中相当高的比例出售给极少一小部分买家，那么库存囤积机器人可能就是罪魁祸首。虽然其中一些机器人会简单地填充并放弃购物车以阻止合法客户，但其他机器人会实际购买您的库存，目的是在其他站点上以更高的价格转售。

客户投诉增加

帐户锁定及欺诈交易相关的支持工单数量上升，可能是凭据填充机器人活动的迹象。通过利用来自过去泄露的信息，这些机器人能够接管合法的用户帐户。这些欺诈性交易不仅会对您的客户体验造成负面影响，还会使服务器超载，从而导致更长的页面加载时间，甚至使您的网站不可用。

登录尝试失败次数增加

每个客户都时不时地输入错误的密码，但是，如果您发现登录尝试失败的次数突然增加，那很可能是遇到了机器人问题。虽然一些凭据填充机器人试图通过盗窃的凭证来访问合法的客户帐户，但一种更简单且更常见的技术是发起暴力攻击，在这种攻击中，机器人使用成千上万的常见用户名和密码词典尝试进行多次快速登录。当机器人超出您站点对特定帐户的登录失败次数限制时，该帐户的真实所有者将被锁定，直到您解决该问题为止，这是一个严重的用户体验问题。

广告支出收效低

数字广告可以有效地为您的网站带来流量，但它对恶意机器人而言也是一个有利可图的武器。很多流量机器人模拟人类用户的行为——反复点击广告来推高您的按点击付费 (PPC) 支出，然后离开页面而不进行任何购买。虽然某些广告平台部署了机器学习算法来减少点击欺诈，很多此类行为依然无法被发现[2]。
因此，必须采取主动，监控广告的每一次点击。

页面浏览量分析失真

如果您的网页浏览量突然上升而没有明显的原因，则可能是恶意机器人在作祟。如果您刚刚推出新产品或活动促销，流量增加可能会来自人类用户，但恶意机器人操作者越来越聪明，会选择在这些时间部署内容抓取机器人，窃取您的内容并对您的总体分析数据产生负面影响。

帐户创建量突增

当成百上千的新用户帐户突然出现时，机器人可能是幕后黑手。恶意机器人可以使用这些虚假的帐户在您的公共评价系统发布垃圾信息，并进行其他多种形式的欺诈，这不仅威胁您的收入和用户留存率，还会威胁您的品牌信誉。

在未经批准的网站上复制您的内容

其他网站分享您的内容是一件好事，但是直接复制的内容突然增加是内容抓取机器人的标志。这些机器人会窃取您花时间收集和整理的信息，让恶意站点运营者将这些内容托管在他们自己的网站上，增加自己的流量，您的网站却遭受损失。

来自异常地理位置的流量

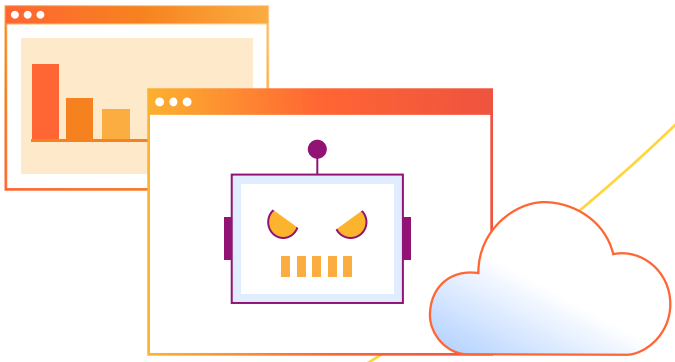
来自意外位置的流量突然增加可能表明有恶意机器人活动，特别是如果这种活动成群出现，且集中在客户没有居住或服务不可用的区域。请密切注意与您的常规用户群无关的任何可疑活动。

非常规时段来自典型位置的流量

就像来自意外位置的流量激增可能指向恶意机器人一样，在异常时间来自正常位置的流量激增也可能表明机器人试图伪装成您的常规用户。例如，如果您在半夜看到某个公共区域的活动量激增，则可能需要更仔细地调查该流量。

信用卡验证失败次数增加

恶意机器人的一个特别危险的迹象是验证失败的信用卡交易数量增加。信用卡填充机器人将测试成千上万个被盗的信用卡号，以试图找到可以利用的信息。恶意机器人通过在安全性较低的网站上进行低价值购买，然后在较大的网站上进行较大交易或在暗网上出售经过验证的卡号来达到目的。您的网站可能会在其中任一环节卷入这些计划中——而且，如果失败交易的情况极其严重，可能会被您的支付服务商处以罚款。



2. https://www.entrepreneur.com/article/313943

对抗机器人的策略

就像没有两个机器人攻击完全一样，您通常需要多种战术的组合才能提前阻止恶意机器人。请考虑下面的部分策略：

一旦发现恶意机器人，立即加以阻止

应对机器人最显而易见的响应也是最有效的措施之一：只需阻止您确定为来自恶意机器人活动的所有流量。仅此一项策略就可以为您节省大量的带宽和存储成本，更不用说维护消费者的信任以及品牌声誉。同时，请记住，如果机器人操作者的动机非常强烈，则可能会改变策略，随后以不同的攻击手段卷土重来。

将所有已知的善意机器人列入允许名单

即使在您检测到和阻止恶意机器人时，确保来自搜索引擎和合作伙伴的善意机器人仍然能够抓取您的网站内容也至关重要。这不仅可以确保您的 SEO 排名保持稳定，还有助于来自第三方服务的合法客户流量顺畅地流入。允许名单还使得设置不会对真实用户访问产生负面影响的机器人阻止规则变得更加容易。

质询检测到的可疑机器人

一旦您发现可疑登录、低页面深度或页面停留时间、信用卡验证失败或任何其他标志性机器人行为模式，部署安全测试至关重要。过去，发送 CAPTCHA 验证码曾被认为是最佳实践。但是，许多先进的机器人现在可以比人类用户更快、更准确地解决这些难题。

如今，最好的方法是使用无 CAPTCHA 验证码的质询，其中“质询”软件会考虑多种因素来确定用户是否实际为机器人（例如，网络、设备、JavaScript 指纹）。
这样的“No CAPTCHA”质询最好与完整的机器人管理解决方案集成，以实现最大的准确性（有关更多信息，请参见结语部分）。

限制用户请求信息的速率

对于较为简单的机器人而言，速率限制是一种有效的手段。通过对任何 IP 地址可以向您的站点提交请求的次数设置硬性限制，可以防止许多简单的暴力机器人攻击（这些攻击试图使用成千上万的词典单词和常用密码进行登录）。但是，当今更先进的机器人可以将其请求数量保持在您的速率限制之下，而在继续造成损害的同时不被发现。

保留所有站点流量的详细日志

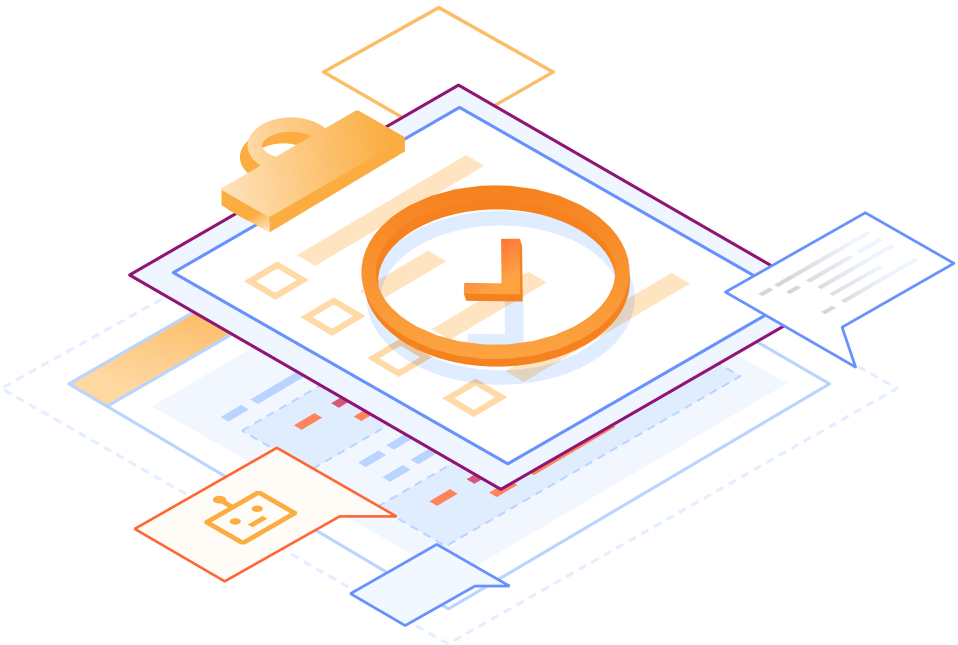
尽管您可能已经维护了每日的页面浏览量和帐户登录日志，但更详细的用户信息日志（例如 IP 地址、浏览器、设备、操作系统、地理位置、引荐来源网址、网络和页面浏览量）对于检测更细微的活动模式而言具有无可估量的价值。日志可以让您清楚地了解机器人在网站上的行为方式，从而使您可以设置更有效的安全策略。此外，在您确实遭受数据泄露的情况下，日志通常对于报告和合规性至关重要。

将机器人重定向到替代内容

当您确定某个流量来源是机器人时，为其提供替代内容，以消耗其计算资源。您甚至可以将伪造的数据（例如错误的定价信息）馈送到内容抓取机器人，使它们对操作者毫无用处。诸如此类的技术将使您有时间观察每个机器人的行为，了解其活动模式，并制定战略以一劳永逸地应对它。

要求所有用户进行附加身份验证

随着机器人攻击的日益普遍，越来越多的站点正在转向加强安全措施，即使对于合法的人类登录也是如此。例如，双因素身份验证 (2FA) 要求用户在多个设备或帐户上确认其身份，而一次性密码 (OTP) 凭据填充书填充机器人。但是请记住，这些技术可能会给登录过程增加摩擦，从而可能会对您的用户体验产生负面影响。

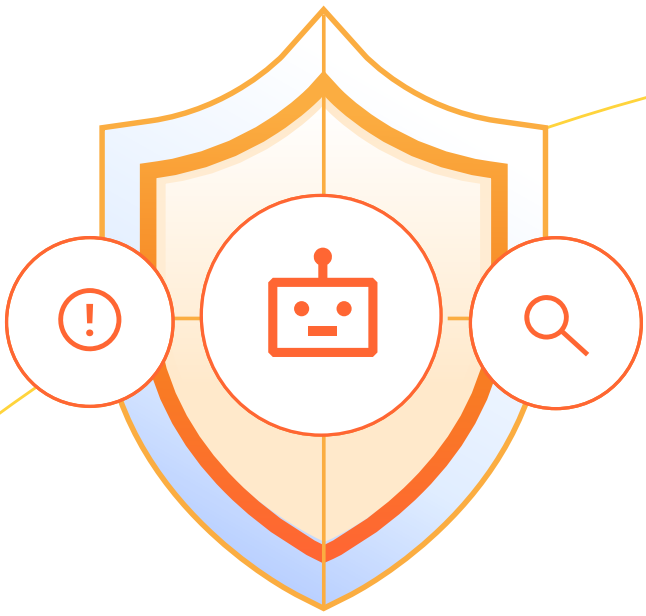


结语

在探索这些策略时，请记住：其中任何一个单独使用都不能阻止所有机器人。为了保护您的业务，您可能需要结合使用各种策略，并添加高级多变量模式分析。Cloudflare Bot Management 可以帮助各行各业的组织采用这种多管齐下的方法。它已整合到整个 Cloudflare 网络中，后者支持超过数以百万计的互联网资产，覆盖全球超过 330 个城市。

通过利用来自整个网络的持续威胁情报，Cloudflare Bot Management 提供行为分析、机器学习和客户端指纹识别，能有效减少对抗恶意机器人所需的精力。此外，Cloudflare 提供 Turnstile，这是一种近乎无感知的、无需 CAPTCHA 的机器人验证方式，只需几行代码即可集成到任何 Web 应用中。

进一步了解 Cloudflare Bot Management。



010 8524 1783 | enterprise@cloudflare.com | www.cloudflare.com/zh-cn/

REV:BDES-6333.2024SEPT03