基于 Web 访问日志的异常行为检测

刘志宏 孙长国
(92117 部队 北京 100072)

[摘 要] 随着互联网的快速发展,各类对网站的攻击技术层出不穷,文章介绍了使用日志分析技术对海量 Web 访问日志进行分析的流程,同时通过使用特征字符匹配、访问频率统计分析等方法去挖掘攻击行为,并通过实际应用场景的展现,描述了在实际攻击发生后如何发现攻击源,从而提高安全威胁的检测能力。

[关键词] 访问日志 日志分析 网络安全
中图分类号: TU712+.3 文献标识码: A 文章编号: 1008-1739(2015)13-62-3

Abnormal Behavior Detection Based on Web Access Log

LIU Zhi-hong, SUN Chang-guo
(Unit 92117 of PLA, Beijing 100072,China)

Abstract: With the rapid development of the Internet, all kinds of site of the attack technology emerge in an endless stream. This paper introduces the use log analysis of huge amount of web access log analysis process, also by using characteristic string matching and access frequency statistical analysis and other methods to excavate the aggressive behavior, through the practical application scenarios to show the described in the actual attack occurred after how to find the source of the attack, so as to improve the detection capability of security threats.

Key words: access log; log analysis; network security

1 引言

随着互联网的快速发展,在经济利益的驱使下,发生了越来越多的网络安全事件,使得互联网安全正在面临前所未有的挑战。攻击者也从单一的攻击行为,发展成有组织、有目标、持续时间长的攻击,新名词 APT 攻击也越来越多的被人们熟悉。攻击者为了获得某个组织甚至国家的重要信息,会利用多种攻击手段,甚至很多攻击利用的漏洞还未公开,在攻击过程中也会使用各种技巧,长期潜伏在系统中,不断收集各种信息,最终达到目的。

对于重要机构和企业而言,虽然已经部署了大量针对某类威胁的安全防护设备,但很多安全设备还是主要依赖于提取漏洞特征的方式检测攻击,如果碰到未公开的漏洞被黑客利用、厂商未及时发布升级包或者管理员未及时更新程序等几种情况,都会导致设备无法检测和防范攻击。为了解决传统安全设备的不足,经过安全技术的积累,考虑通过 Web 访问日志分析的手段,对 Web 用户的行为进行异常监控,发现可疑行为,从而进一步追查原因,及时采取措施应对问题。

定稿日期:2015-06-12

2 日志分析流程

常见日志分析流程见图 1[1]:

图 1 日志分析流程

(1)日志采集

Web 访问日志有两种采集方式,一种是直接导入 Web 服务器的访问日志,另一种是从 Web 应用防火墙的访问日志接口导入,导入的是经过 Web 应用防火墙清洗后的 HTTP 数据。

(2)日志入库

采集到的 Web 访问日志将根据字段进行切割,保存到数据库中等待分析。

(3)数据分析

主要通过前面介绍的几种分析方法,编写算法和规则,进行海量数据分析,然后报告可疑行为。

(4)结果展现

将分析的可疑行为,以图表、地图定位等可视化的方式展现给安全管理员。

3 访问日志结构

访问日志是 Web 服务器(例如 Apache、IIS)记录用户访问行为产生的文件,标准的 Web 日志是纯文本格式,每行一条记录,对应客户端浏览器对服务器资源的一次访问。典型的日志包括来源地址、访问日期、访问时间、访问 URL 等丰富的信息,对日志数据进行分析,不仅可以检测到可疑的漏洞攻击行为,还可以提取特定时间段、特定 IP 对应用的访问行为[2]。

由于不同 Web 服务器的日志格式不同,以 Apache 访问日志为例,介绍日志组成字段及字段的意义,如图 2 所示[3]。

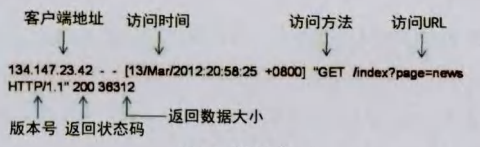

图 2 字段含义

上述日志记录了来自 134.147.23.42 的地址,在北京时间 2012 年 3 月 13 日 20 点 58 分 25 秒,对服务器的 /index?page=news 页面进行了一次访问,访问使用 GET 方式,使用 HTTP 1.1 协议,返回状态码是 200(页面存在,请求成功),返回的页面大小是 36312 字节。

Apache 日志格式可以通过编辑 httpd.conf 文件自定义,默认有简单和复杂两种格式,见图 3、图 4:

Common Log Format

```
LogFormat "%h %l %u %t \"%r\" %>s %b" common
CustomLog logs/access_log common
```

图 3 简单格式

Combined Log Format

```
LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-agent}i\"" combined
CustomLog log/access_log combined
```

图 4 复杂格式

复杂格式比简单格式多两个字段:referer 和 user-agent。也可以自定义格式,各字段含义如表 1 所示:

表 1 字段含义

字段	描述
%h	客户端地址
%l	Identd登录名
%u	HTTP认证用户名
%t	访问时间
%r	访问方法、访问URL、协议版本
%>s	返回状态
%b	返回大小
%{Referer}i	来源站点

4 常见分析方法

4.1 特征字符分析(Signature-based)

此方法是在日志中查找已知的漏洞特征,去发现黑客攻击行为,是最简单的方法。

例如,可以通过匹配扫描器请求 URL 中的特征,检测漏洞扫描行为;通过匹配 URL 中的 SQL 关键词,检测 SQL 注入攻击。

4.2 访问频率分析(Frequency analysis)

在黑客攻击过程中,需要对系统进行各种特定的访问,这些访问与正常用户访问有很大差别,每种攻击行为都有不同的特征。通过对大量用户访问数据的挖掘,可以发现这些异常访问行为。

5 应用场景展现

2013 年某天接到用户电话,描述在某论坛发现有网友发布了大量用户数据,初步判断可能是某业务系统被黑客非法登录,然后批量下载了用户数据。工程师到达现场后,查看了防火墙、IDS、Web 应用防火墙日志,均未发现攻击记录,又对 Web 应用进行了安全检测,也未发现明显的 Web 漏洞。跟业务人员沟通后得知未使用 123456 等弱口令。初步检查未果后,工程师将客户 Web 服务器的访问日志导入数据分析平台,进行分析后发现,大约在 1 周前,周三至周五每天晚上都有大量的请求后台登录 URL 的访问记录,系统对请求的源 IP 进行归并、提取分析,最终定位了 1 个成功登录过系统的最可疑的 IP,并且归属地是某地。

将此 IP 的访问记录筛选出来,分析梳理出攻击流程:

获得管理员邮箱地址 -> 查询管理员密码 -> 用字典破解出后台密码 -> 用密码登录到系统后台。

6 结束语

本文中提出了使用 Web 访问日志分析的方法检测异常行为，可以利用此方法实现一个 Web 访问日志分析系统，配合传统 Web 应用防火墙进行安全防范。实际使用中，外部用户的访问首先到达 Web 应用防火墙，由 Web 应用防火墙进行检测、过滤，干净的流量进入 Web 服务器。通过对 Web 访问日志分析，不仅能补充检测 Web 应用防火墙没有检测到的攻击，也可以在出现新漏洞后，及时统计哪些 Web 应用受到影响，从而切实提高网络威胁检测和防御能力。

参考文献

[1]钱秀槟，李锦川，方星.信息安全事件定位中的 Web 日志分析方法[J].信息网络安全，2010(06)：59-61.

[2]李雪. 基于大数据实时 web 防火墙日志安全审计系统的探究[J].网络安全技术与应用，2014(12)：109-110.

[3]张峰，付俊，杨光华. Web 访问日志安全分析技术研究[J].北京邮电大学学报，2014(2)：93-95.

卡巴斯基揭秘神秘 APT 攻击现形记

近日，卡巴斯基实验室宣布，公司检测出一种利用多达三种零日漏洞的最新恶意软件平台——Duqu 2.0。而该恶意软件平台是迄今为止发现的技术最先进、最神秘同时也是最强大的高级可持续性威胁(APT)之一。Duqu 2.0 攻击包括多个未曾发现的独特功能，而且几乎不会留下任何痕迹，致使其很难被检出。尽管 Duqu 2.0 攻击组织的思维理念与方式遥遥领先于其他 APT 攻击组织，卡巴斯基实验室所拥有的领先技术与顶级研究专家使公司成功发现了该攻击组织。目前，卡巴斯基实验室的产品已将这种威胁检测为 HEUR:Trojan.Win32.Duqu2.gen，并且能够高效保护企业与个人用户免受 Duqu 2.0 侵扰。

实际上，早在 2015 年初的一次测试中，卡巴斯基专家就发现了这一威胁。当时，公司正在开发一款 APT 防御解决方案。正是该方案的原型显示出公司网络遭遇复杂针对性攻击的迹象。此后，卡巴斯基启动了内部调查。由公司研究人员、逆向工程师和恶意软件分析师组成的专业团队深入分析了这种独特的攻击，并最终发现了 Duqu 2.0。

卡巴斯基安全专家的分析显示，这是一起精心策划和实施的网络间谍行动，其幕后黑手就是 2011 年臭名昭著的 Duqu 背后组织。卡巴斯基认为，该攻击行动受到了政府的支持。攻击者的主要目的在于监视卡巴斯基的技术、最新研究以及内部流程。除了试图盗窃公司的知识产权(包括卡巴斯基的安全操作系统、卡巴斯基反欺诈解决方案、卡巴斯基安全网络和 APT 防御解决方案以及服务)和高级针对性攻击的近期研究数据外，此次攻击并未出现其他恶意行为，因此不会对公司的产品、技术和服务造成影响。目前，卡巴斯基已采取有效措施确保公司用户及其合作伙伴处于非常安全的状态，并且避免类似的问题再次发生。而通过获取到的攻击信息，卡巴基斯将进一步提升其多款产品的性能。

然而，卡巴斯基并非这种强大的网络攻击的唯一目标。卡巴斯基研究专家还在西方国家、中东和亚洲地区发现了其他受害者。尤其值得注意的是，发生于 2014 至 2015 年的一些最新感染同 P5+1（伊朗核问题六方会谈）会议及其地点有关。Duqu 幕后的攻击者似乎在会议地点发起了攻击。除了 P5+1 会谈，攻击者还针对纪念奥斯维辛集中营解放 70 周年相关仪式和会议发动了类似的攻击。该活动的与会者多为外国政府高官和要员。

在谈及这一重大发现时，卡巴斯基实验室 CEO 尤金·卡巴斯基表示："监视网络安全公司是一件非常危险的事情。当今世界，硬件和网络设备均可能被攻陷，而安全软件则是保护企业和消费者的最后一道防线。而且，恐怖分子和专业网络罪犯早晚会发现并利用这些应用于类似针对性攻击中的技术。这是一种很可能会发生的严重情况。"

对于如何应对此类攻击，尤金·卡巴斯基有着独到的见解："让世界更加安全的唯一方法便是公开此类攻击事件。这样做有助于提高和改进企业基础设施的安全设计，并且向恶意软件开发者发出明确的信号，即所有的非法行为都会被制止，责任人将受到惩罚。唯有执法机关和安全企业公开携手对抗此类攻击，方能保护世界安全。因此，我们会公开任何攻击，不管其源自何处。"

根据卡巴斯基所掌握的情况判断，出于类似的地缘政治利益，Duqu 2.0 还被用于攻击一些高层目标。为清除这一威胁，卡巴斯基实验室进行了初步安全审计和分析，通过 Securelist 公开发布了有关 Duqu 2.0 的所有技术详情，并且愿意向所有感兴趣 / 受此影响的组织提供帮助。目前，审计仍在进行中，并将于几周后完成。此外，卡巴斯基实验室已经联系了多个国家的网络安全组织部门，正式要求对这次攻击进行刑事调查。