安全小课堂第108期【web安全日志分析】

京东安全应急响应中心 8月27日

web日志作为web服务器重要的组成部分，详细地记录了服务器运行期间客户端对web应用的访问请求和服务器的运行状态。同样，攻击者对网站的入侵行为也会被记录到web日志中。因此，在网站日常运营和安全应急响应过程中，我们可以通过分析web日志并结合其他一些情况来跟踪攻击者，还原攻击过程。

JSRC **安全小课堂第108期**，邀请到**bingo**作为讲师就**web安全日志分析**为大家进行分享。同时感谢朋友们的精彩讨论。

为什么需要对日志进行分析?

京安小妹

bingo:

日志文件是在服务或应用程序运行时期间发生事件和操作的关键信息记录，例如：在什么时间由谁以什么方式访问了服务器什么资源。这些信息为性能监控、故障排查、程序调试、调查取证、攻击事件调查提供极有价值的信息。

几乎所有服务器，服务和应用程序都提供某种日志记录。鉴于本课主要讲web日志，如果先看看web服务器的日志例子，下面是由IIS服务器记录的一条POST访问日志。

```
2017-09-18  13:56:48 10.160.23.22 POST /backdoor.aspx - 80 - 220.178.79.254 Mozilla/5.0+(Windows+NT+6.1;+WOW64)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/49.0.2623.221+Safari/537.36+SE+2.X+MetaSr+1.0  200 0 0 109
```

该条日志记录了：在2017-09-18 13:56:48由IP地址为 220.178.79.254，浏览器User-Agent为Mozilla/5.0+(Windows+NT+6.1;+WOW64)+AppleWebKit/537.36+(KHTML,+like+Gecko)+Chrome/49.0.2623.221+Safari/537.36+SE+2.X+MetaSr+1.0的访问者发起了对backdoor.aspx的POST请求，访问成功(状态码：200)并返回 109 个字节的数据。

所有用户对网站的访问都以上述方式保存到了日志文件中。web服务器的日志记录功能默认是开启的，在临时的应急响应中，从web日志分析攻击行为是非常有必要的，可以通过时间、IP、User-Agent、访问路径等特点（俗称用户指纹）还原相关事件经过。在缺失这些日志时，基本就算是凉了。

So，在网站被黑之后的应急响应中，日志分析绝对是必不可少的工作。

讲师

WEB日志安全分析原理?

bingo：

当我们处理应急响应事件时，我们首先需要了解的是网站遭受了什么攻击？并确定可疑攻击者特性(指纹)，如IP地址、UA信息、时间段等

网站运维发现的异常

对于事件应急来讲，客户基本都可以描述清楚大致情况，比如网站出现黑页、后门文件、数据泄露、大量用户投诉等等。一般我们可以直接进行针对性的日志分析。

- 数值统计

在客户的描述信息不足以确定攻击类似的情况下，可以考虑使用数值统计手段来初步识别可能的攻击。如：

- IP访问数量统计 ： 网站扫描
- URL访问数量统计：暴力破解(高统计值)、网站遍历(低统计值)
- 非200请求比：网站遍历、报错注入
- 。。。

【关键字】

- SQL注入 ： select、union、sleep等
- XSS ： alert(、script、src=http等
- 命令执行：whoami、ipconfig/ifconfig、net user等
- 目录遍历 ： ../../等
- 其它 ： dese64_decode、OgnlContext等

【相关日志筛选】

筛选出攻击相关的日志记录，缩小排查范围。

- 指定IP筛选
- 指定UA筛选
- 指定URL筛选
- 组合筛选条件

【攻击类型分析】

在缩小后的范围内，大致查阅记录，一般就能确定可能的攻击类型。 也可以考虑使用关键字（或正则表达式）搜索。

【攻击细节分析】

查看相关日志记录的始末时间，确定是否需要扩大日志范围。对攻击范围内的日志进行分析（也就是查看URL及参数，相对容易能看明白其攻击类型），梳理攻击路径及时间线，以便评估损失、确定漏洞也是为后续报告提供材料。

讲师

如何进行日志分析?

京安小妹

bingo:

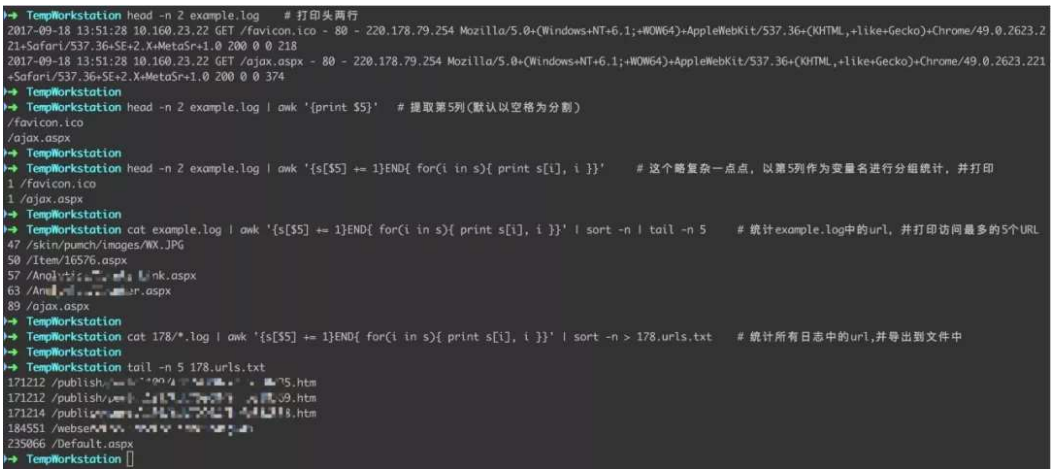
【基本功】

- 命令行工具

掌握grep/awk/sort等命令的基本使用，比如统计URL访问数量：

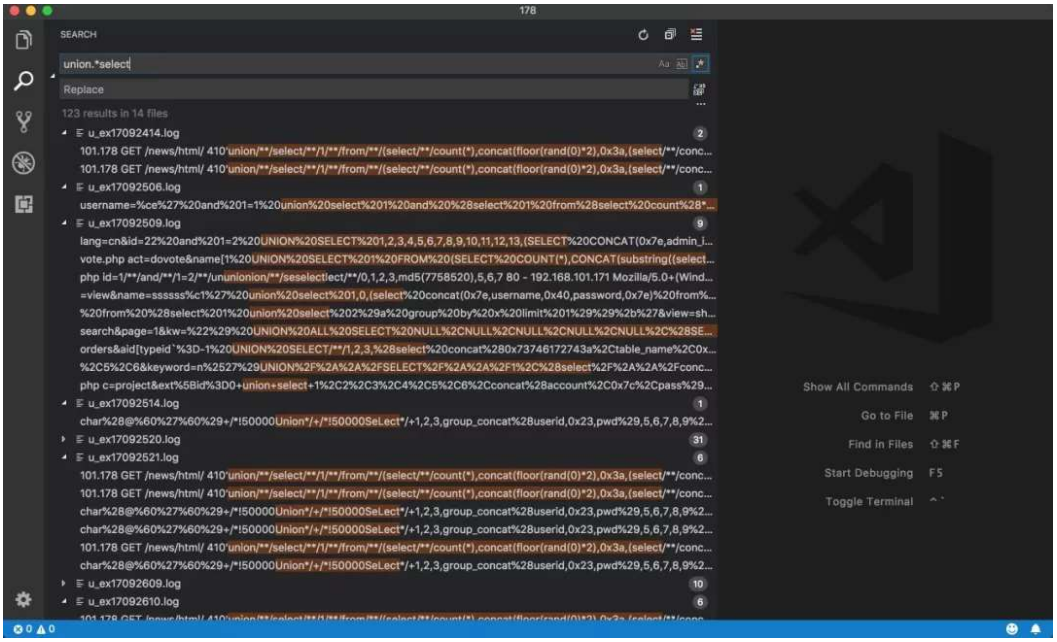
```
cat *.log | awk '{print $5}'| awk '{s[$1] += 1}END{ for(i in s){  print s[i], i } }' | sort -n > urls.txt
```

稍作演示：



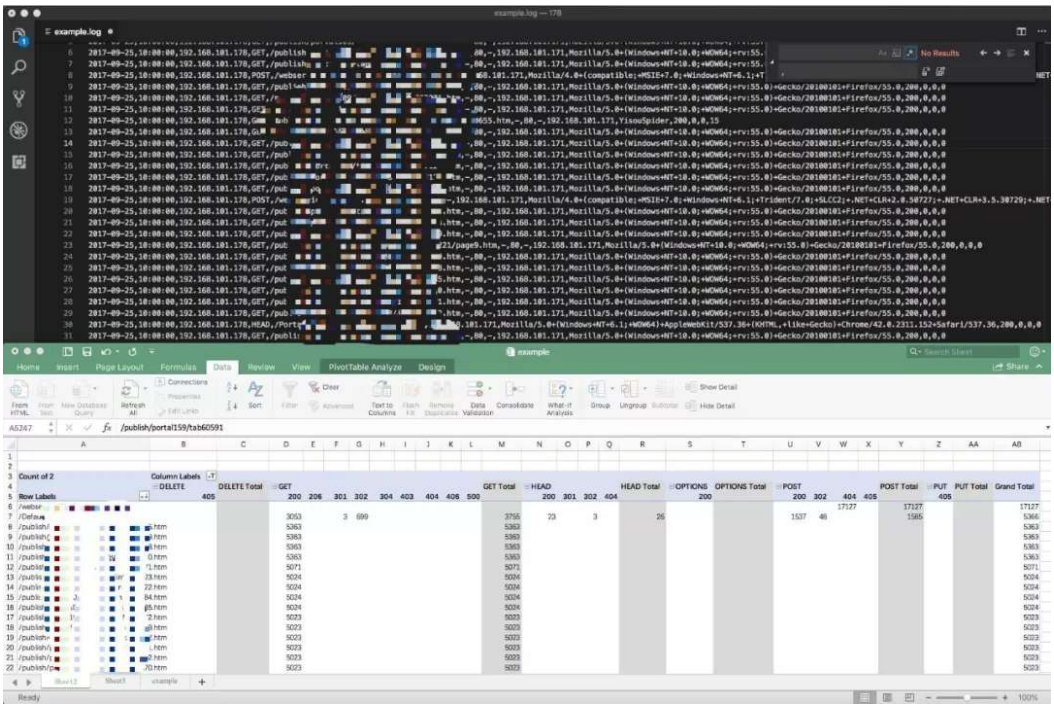
- 文本编辑器 + Excel

使用VScode之类的文本编辑器，一次加入所有相关日志，对所有日志记录进行批量查找、快速定位可能存在的攻击。



如果涉及的记录不是太多（百万条以内），可以将日志的空格替换为逗号并保存为csv文件，然后直接用Excel打开，再使用Excel的批量统计功能进行直观处理。如果出现导入Excel时，存在字段不对应的情况(User-Agent信息中可能包含空格)，可以在Excel中进行批量处理，或者在文本编辑器中使用正则表达式替换避免空格带来的影响。

大致情况如下：



【各种日志工具】

各种专门的工具比起手动来讲会更高效一些，如果觉得有必要可以试用一下，这里就不深入讲解了。

- 360星图 http://wangzhan.360.com/activity/xingtu
- 安全易 https://www.anquanyi.com

讲师

日志分析中存在的难题有哪些？

京安小妹

bingo：

- 日志不记录POST详细参数

日志中POST数据是不记录详细参数的，所以攻击者如果找到的漏洞点为POST请求，那么日志中只能看到POST请求，而不能看到具体提交的参数。

- IP地址不一定可靠

攻击者可能使用多个代理IP，假如我是一个恶意攻击者，为了避免日后攻击被溯源、IP被定位，会使用大量的代理IP从而增加分析的难度，如果一个攻击者使用了大量不同的IP进行攻击，那么使用上面的方法可能就无法进行攻击行为溯源了。

- User-Agent不可靠

与IP地址同理，但比IP地址更不靠谱，只要攻击者愿意，可以每个请求切换一次。

- APT攻击

攻击者分不同时间段进行踩点、攻击，导致时间上无法对应出整个攻击行为，而且这种攻击模式下，攻击者基本就是隐匿IP地址的，基本难成狗。

- 其它

其它的一些情况不一而足。比如：如果服务器前端加了负载均衡，日志里面记录的就不是访问者的IP了（而是负载均衡的IP）；具体攻击中，可能攻击者一边跑着扫描器，一边手工渗透，将导致大量"噪音"记录，严重干扰攻击路径的梳理。

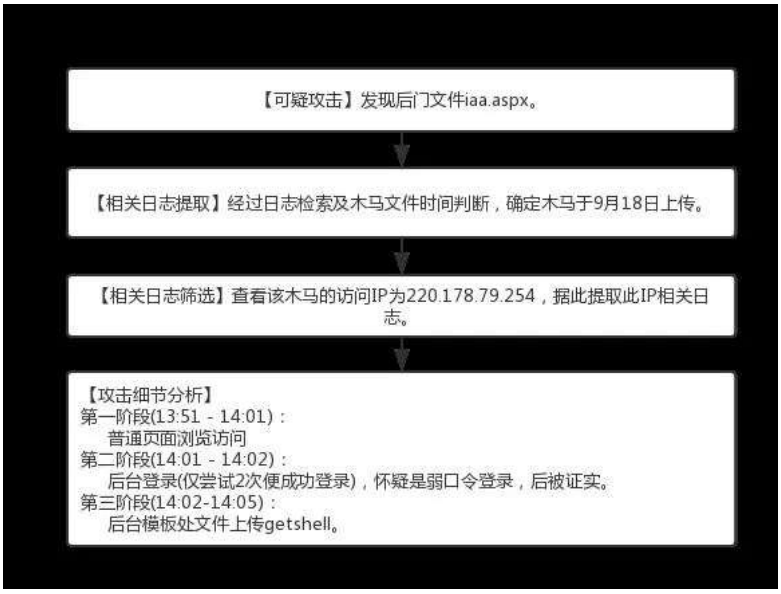
讲师

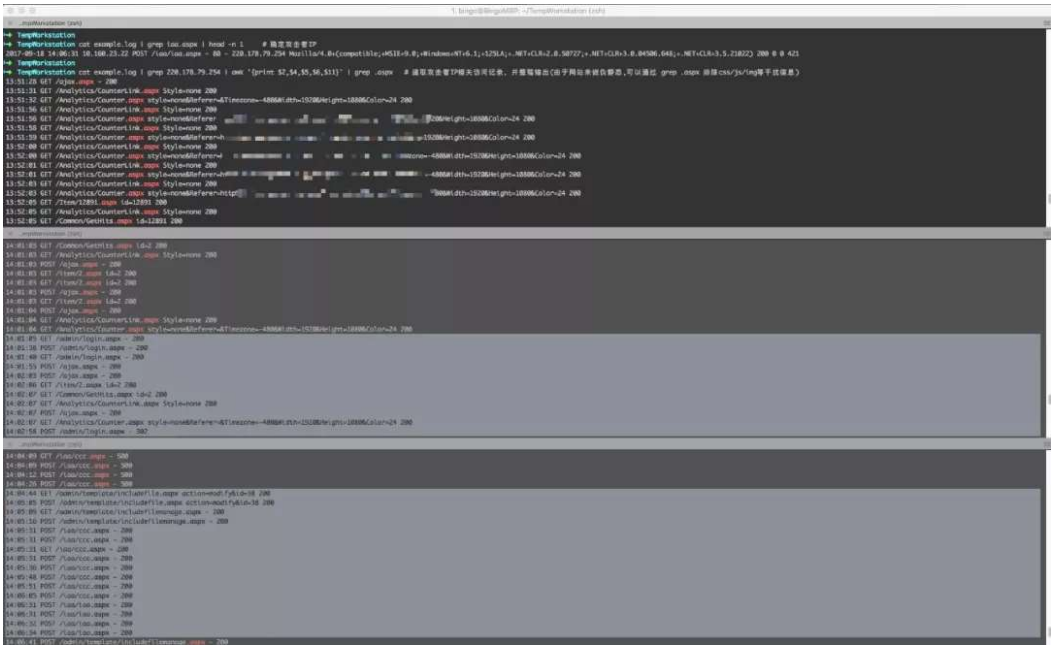
web日志实例分析

京安小妹

bingo：

某网站发现存在后门，故响应客户需求，进行相关排查。下面只说排除中日志相关的部分，按照咱们之前提到的排查思路一步一步来：





讲师

互动问答环节：

1. NGINX，有什么方法可以记录到post数据到log里面?

讲师:

默认情况下都是不记录POST数据包的，在应急响应的情况下，不对此做奢求。

2. 免费的win日志分析有什么?

讲师:

LogParser、360星图、安全易可以看看

3. 应急响应中，日志被删了怎么搞?

讲师:

这个就涉及到数据取证了。

一般的，需要对整个磁盘做克隆（每个bit都保存下来），然后做数据还原。看是否有机会还原日志数据

4. 有什么办法解决日志存储时间?

讲师:

日志时间一般不太有问题的，如果有时区问题，那么对应的进行时间加减就行了。 除非攻击者人为对日志进行了修改伪造。

一些中间件是可以设定的，不过你得考虑你的硬盘够不够用,具体的设定方法可以参考各服务的文档。

5. 针对于病毒的应急可以讲讲么?

讲师:

看它的危害程度，如果有内网扩散的风险，那至少先做一定的网络隔离；如果是有通信连接的话，建议保存内存镜像以便后续对内存内容进行分析。同时提取病毒样本，然后回实验室找反汇编、反编译大牛来分析。

6. 方便提供几个日志分析工具?

讲师:

logParser awk/grep

日志分析还是要看用途，很多的日志分析工具并不是针对特定的安全事件，太多的冗余信息。

对于特定的事件处理来讲，没有一针见血的效果。

本期JSRC 安全小课堂到此结束。更多内容请期待下期安全小课堂。如果还有你希望出现在安全小课堂内容暂时未出现，也欢迎留言告诉我们。

安全小课堂的往期内容开通了自助查询，点击菜单栏进入“安全小课堂”即可浏览。

简历请发送：cv-security@jd.com

微信公众号：jsrc_team

新浪官方微博：京东安全应急响应中心