Web日志安全分析浅谈

jeary / 2017-06-14 03:32:00 / 浏览数 42563 技术文章 技术文章 顶(8) 踩(0)

作者： jeary@安百科技

一、为什么需要对日志进行分析？

随着Web技术不断发展，Web被应用得越来越广泛，所谓有价值的地方就有江湖，网站被恶意黑客攻击的频率和网站的价值一般成正比趋势，即使网站价值相对较小，也会

此时对网站的日志分析就显得特别重要，作为网站管理运维等人员如不能实时的了解服务器的安全状况，则必定会成为"被黑了还不知道的"那一类人，从而造成损失，当然还

二、如何进行日志分析？

在说如何进行分析之前，我们先来了解一下Web服务器中产生的日志是什么样子.
我们以Nginx容器为例：

随机抽取一条日志：

```
61.144.119.65 - - [29/May/2017:22:01:32 +0800] "GET /page/1 HTTP/1.1" 200 6403 "http://www.baidu.com" "Scrapy/1.1.2 (+http://s
```

作为Web开发或者运维人员，可能对图中的日志信息比较熟悉，如果对日志不那么熟悉也没关系，我们可以查看Nginx中关于日志格式的配置，查看nginx.conf配置文件：

可以看到日志格式为：

```
$remote_addr - $remote_user [$time_local] "$request" '$status $body_bytes_sent "$http_referer" '$http_user_agent" "$http_x_for
```

翻译过来即为：

```
■■IP - ■■■■■  ■■■■■■ ■■■■■ ■■■■■ ■■■■■■ ■■■■■ ■■■■■■ ■■■■■■IP
```

通过以上信息，我们可以得知服务器会记录来自客户端的每一个请求，其中有大量来自正常用户的请求，当然也包括来自恶意攻击者的请求，那么我们如何区分正常请求和恶

(图中以"select"为关键字进行过滤)
聪明的你肯定想到了，如果此时加上时间条件，状态码等条件就能查询到最近可能成功的SQL注入攻击了，当然实际情况中，仅仅只依靠状态码来判断攻击是否成功是不可行

```
/logo.png?attack=test';select/**/1/**/from/**/1
```

此时请求状态码为200，但是此注入攻击并没有得到执行，实际情况中，还会有更多情况导致产生此类的噪声数据。
抛开这类情况不谈，我们来说说在一般应急响应场景中我们分析日志的常规办法。
在常规应急响应常见中，一般客户会有这几种被黑情况：

1.带宽被占满，导致网站响应速度变慢，用户无法正常访问
2.造成已知经济损失，客户被恶意转账、对账发现金额无端流失
3.网站被篡改或者添加暗链，常见为黑客黑页、博彩链接等
..
对于这些情况，按照经验，我们会先建议对已知被黑的服务器进行断网，然后开始进行日志分析操作。假设我们面对的是一个相对初级的黑客，一般我们直接到服务器检查是
1.搜索最近一周被创建、更新的脚本文件
2.根据网站所用语言，搜索对应webshell文件常见的关键字
找到webshell后门文件后，通过查看日志中谁访问了webshell，然后得出攻击者IP，再通过IP提取出攻击者所有请求进行分析
如果不出意外，可能我们得到类似这样一个日志结果：（为清晰呈现攻击路径，此日志为人工撰造）
eg:

```
00:01  GET http://localhost/index.php 9.9.9.9  200  [■■■■■]
00:02  GET http://localhost/index.php?id=1' 9.9.9.9 500  [■■■■■]
00:05  GET http://localhost/index.php?id=1' and 1=user() or ''=' 9.9.9.9  500  [■■■■■]
00:07 GET http://localhost/index.php?id=1' and 1=(select top 1 name from userinfo) or ''=' 9.9.9.9 500 [■■■■■]
00:09 GET http://localhost/index.php?id=1' and 1=(select top 1 pass from userinfo) or ''=' 9.9.9.9 500 [■■■■■]
00:10  GET http://localhost/admin/ 9.9.9.9 404 [■■■■■]
00:12  GET http://localhost/login.php 9.9.9.9 404 [■■■■■]
00:13  GET http://localhost/admin.php 9.9.9.9 404 [■■■■■]
00:14  GET http://localhost/manager/ 9.9.9.9  404 [■■■■■]
00:15  GET http://localhost/admin_login.php 9.9.9.9 404 [■■■■■]
00:15  GET http://localhost/guanli/ 9.9.9.9 200 [■■■■■]
00:18  POST http://localhost/guanli/ 9.9.9.9 200 [■■■■■]
00:20  GET http://localhost/main.php 9.9.9.9 200 [■■■■■]
00:20  POST http://localhost/upload.php 9.9.9.9 200 [■■■■■]
00:23  POST http://localhost/webshell.php 9.9.9.9 200 [■■■■■]
```

```
00:25  POST http://localhost/webshell.php 9.9.9.9 200 [■■■■]
00:26  POST http://localhost/webshell.php 9.9.9.9 200 [■■■■]
```

首先我们通过找到后门文件“webshell.php”，得知攻击者IP为9.9.9.9，然后提取了此IP所有请求，从这些请求可以清楚看出攻击者从00:01访问网站首页，然后使用了单引号
从以上分析我们可以得出，/index.php这个页面存在SQL注入漏洞，后台地址为/guanli.php,/upload.php可直接上传webshell
那么很容易就能得出补救方法，修复注入漏洞、更改管理员密码、对文件上传进行限制、限制上传目录的执行权限、删除webshell。

三、日志分析中存在的难题

看完上一节可能大家会觉得原来日志分析这么简单，不过熟悉Web安全的人可能会知道，关于日志的安全分析如果真有如此简单那就太轻松了。其实实际情况中的日志分析
对于日志的安全分析，可能会有如下几个问题，不知道各位可否想过。
1.日志中POST数据是不记录的，所以攻击者如果找到的漏洞点为POST请求，那么刚刚上面的注入请求就不会在日志中体现
2.状态码虽然表示了响应状态，但是存在多种不可信情况，如服务器配置自定义状态码。
如在我经验中，客户服务器配置网站应用所有页面状态码皆为200，用页面内容来决定响应,或者说服务器配置了302跳转，用302到一个内容为“不存在页面”（你可以尝试用
3.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（淘宝上，一万代理IP才不到10块
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
4.无恶意webshell访问记录，刚才我们采用的方法是通过“webshell”这个文件名从日志中找到恶意行为，如果分析过程中我们没有找到这么一个恶意webshell访问，又该从
5.分析过程中我们还使用恶意行为关键字来对日志进行匹配，假设攻击者避开了我们的关键字进行攻击？比如使用了各种编码，16进制、Base64等等编码，再加上攻击者使
6.APT攻击，攻击者分不同时间段进行攻击，导致时间上无法对应出整个攻击行为
7.日志数据噪声（~~这词我也不知道用得对不对~~）上文提到过，攻击者可能会使用扫描器进行大量的扫描，此时日志中存在大量扫描行为，此类行为同样会被恶意行为关键字匹
..

四、日志分析工程化之路 [探索篇]

（上一节留下的坑我们留到最后讨论[~~因为我也觉得比较头疼~~]，我们现在来讨论一点让人轻松的~）
曾经有运维的人员问我们公司的大神，该如何分析日志？
大神回答了三个字：“用命令”
因为站在安全经验丰富的人角度来看，的确用命令足矣，可是对于安全经验不那么丰富的人来说，可能就不知道从何入手了。但是即使身为一个安全从业人员，我也觉得用命
于是，聪明的黑客们就想到了，将这些步骤流程写成工具，让工具来帮我们分析日志，当然我也想到了，可是在我造这么一个轮子之前，我习惯性的到各大网站上先翻一翻，

腾讯安全实验室：
https://security.tencent.com/index.php/opensource/detail/15

北风飘然@金乌网络安全实验室
http://www.freebuf.com/sectool/126698.html

网络ID为piaox的安全从业人员：
http://www.freebuf.com/sectool/110644.html

网络ID：SecSky
http://www.freebuf.com/sectool/8982.html

网络ID：鬼魅羊羔
http://www.freebuf.com/articles/web/96675.html

我以“Web安全日志分析”为关键字，百度&Google了一番，发现并没有找到自己觉得不错的日志分析工具，难道安全行业就没有大牛写个优秀的日志分析工具出来？年轻时

首先是推广做得比较好的:日志易
https://www.rizhiyi.com/

日志易确实像它推广视频里所说的：“国内领先的海量日志搜索分析产品”
前段时间，有客户联系到我们，说他们买了日志易的产品，但是其中对安全的监控比较缺乏，让我们能不能在日志易的基础上添加一些安全规则，建立安全告警，他们要投放
但是日志易确实有几个优点
1.日志采集方面相对成熟，已经能针对多种日志格式解析并结构化，还支持用户自定义日志格的辅助解析
2.海量日志存储相对完善，可接收来自各个客户端的日志，Saas服务成熟，能对接各大云主机
3.搜索方面技术优秀，千亿级别数据索引只需60秒
（但是，我要的安全分析啊，其他的再成熟，也始终是个不错的日志分析平台而已，我要的是安全分析、安全分析、安全分析[重要的话说三遍]）
补：
（后来我发现，日志易其实有在安全方面进行分析，但是这个如图这个结果，并没有让我觉得眼前一亮，而且其中还有大量的误报）

后来我从朋友那里得知另外一个产品，算是看到一个稍微像那么回事的产品：
安全易
https://www.anquanyi.com/

他们推广做得不那么好，所以在我一开始的搜索中，并没有从搜索引擎找到它，这个产品是可以免费注册并试用的，于是我迫不及待注册了一个账号进去看看，如图：

当我试用过安全易这个产品之后，提取出了他们在关于安全方面所做的统计列表，如下：

1.威胁时序图
2.疑似威胁分析
3.疑似威胁漏报分析

4.威胁访问流量
5.威胁流量占比
6.境外威胁来源国家(地区)统计
7.境内威胁来源城市统计
8.威胁严重度
9.威胁响应分析
10.恶意IP
11.恶意URL分析
12.威胁类型分析
13.威胁类型分布
14.威胁分类计数
15.威胁来源热力图
16.威胁总数
17.威胁日志占比
结果似乎挺丰富，至少比我们开始使用命令和工具得到的结果更为丰富，其实在看到这个产品之前，我们内部就尝试使用过各种方法实现过其中大部分视图结果，但是似乎这
——
攻击行为溯源，也就是我们在第二节中对日志进行简单的分析的过程，得到攻击者的整个攻击路径已经攻击者执行的恶意操作。不过想要将这个过程工程化，难度可比如上1
虽然安全易的产品并没有满足我对日志分析中的想法，但是也不能说它毫无价值，相反这款产品能辅助运维人员更有效率的监控、检查服务器上的安全事件，甚至他们不用懂

五、日志分析工程化之路 [实践篇]

在了解了很多分析日志的工具后，也尝试过自己折腾出一个方便分析日志的工具，以便以日常工作中的应急响应场景
记得是在半年前左右，我的思路是这样的：

1.首先确认日志结构
我在Mysql中建立了如下结构的一张表来存储日志：

```
■■■■
   ■■■■
   ■■■■■
   ■■■IP
   ■■■■
   ■■■■
   ■■■■■
   ■■■IP
   ■■■■■
   ■■■■■
   ■■■■
   ■■■■
   ■■■■
```

2.给Web攻击进行分类

```
■■■■■
   ■■■■■■
   ■■■■
  ■■/■■
```

3.建立攻击规则表对应不同的攻击类型

```
■■■■■
   ■■■■■■■■■
   ■■■■■■■■
   ■■■■■■■■■■■■ID
```

此时不得不说一下当时日志是怎么入库的，不知道大家是否知道awk命令

```
echo "aa bb cc" | awk -F '{print $1}'
```

我们对日志采用了类似的方式，通过空格分割，然后生成出Mysql中可用的insert语句
大约为：`INSERT INTO web_log VALUES ($1,$3,$4,...)`,至于你说其中列数是如何对应到Mysql里的表结构的,我们当时是人工核对的，为每一个不同的日志文件进行人工对应..（可想而知这活工作量多大）
扯回正题，当我们入库完毕后有了这么三张数据表，聪明的童鞋可能已经知道下一步要干什么的，那就是拿着安全规则正则表达式去逐条匹配日志表里面的日志
然后得到结果：

```
■■■■ID
■■■■ID
■■■■ID
```

最后我们只需要写SQL语句，就能轻松统计各个攻击类型都分别有多少攻击请求了
如图：

最后我们思考了从各个角度来进行查询，得到了如下以下列表中的结果：

网站受攻击次数排名
网站高危请求排名
网站攻击者数量排名
网站受攻击页面排名
可疑文件排行.
被攻击时间统计
攻击来源分布
高危攻击者排行
攻击者攻击次数排行
网站危险系数排行
攻击者数量统计
各站点攻击者数量统计
各扫描器占比
使用扫描器攻击者统计
由于这是一次针对多个(70+)站点的分析，所以只需突显安全趋势即可，在此次日志分析中，还并未涉及到溯源取证
记得当时我们是用SQL语句查询出结果，然后将数据填入Execl，然后进行图标生成，整个日志分析的过程，从日志原文件到入库到匹配到统计到出数据最后到Execl出统计图
对了，上几张图瞧瞧吧

（篇幅原因，其他统计图不贴上来了）
可以看出，我们仅仅只是采用了一些安全攻击规则来对日志进行匹配就已经得到了不错的结果，虽然整个过程有点漫长，但是得到的这一份日志分析报告是具有实际意义和价
~~但是日志分析工(lan)程(duo)化(hua)之路到此结束了吗？不，远远没有~~

六、日志分析工程化之路 [进阶篇]

有没有觉得像上面那样折腾太累了，虽然确实能得到一个还不错的结果。于是开始找寻一个较好的日志分析方案，最后采用了开源日志分析平台ELK,ELK分别为：

Elasticsearch 开源分布式搜索引擎
Logstash 对日志进行收集、过滤并存储到Elasticsearch或其他数据库
Kibana 对日志分析友好的Web界面,可对Elasticsearch中的数据进行汇总、分析、查询
因为它开源、免费、高可配，所以是很多初创企业作为日志分析使用率最高的日志分析平台
当理清楚ELK的搭建方式和使用流程后，我们用ELK实现了一遍第五节中的日志分析
流程大概如下：
1.编写Logstash配置文件

2.将攻击规则应用于logstash的filter插件

3.利用载入了安全分析插件后的logstash进行日志导入

4.查询分析结果

5.利用Kibana进行统计、可视化

到这里，所得结果已经比"HanSight瀚思"安全易这个产品的结果更为丰富了~
，但是日志安全分析之路远远没有结束，最重要也最具有价值的那部分还没有得到实现 —— 攻击行为溯源

七、日志安全分析攻击溯源之路 [探索篇]

故技重施，我搜寻了和攻击溯源有关的相关信息，发现国内基本寥寥无几
最后发现其实现难度较大，倒是听说过某些甲方内部安全团队有尝试实现过，但至今未要到产品实现的效果图，不过最后倒是被我找到某安全公司有一个类似的产品，虽然

由于图1中的分析已经实现，这里暂且不谈。我们看图2中的攻击溯源，这好像正是我们需要的效果。
第一个信息不难理解，三个中国的IP发起了含有攻击特征的请求，他们的客户端信息(userAgent)分别为Linux/Win7/MacOs
第二个信息据我经验应该是他们内部有一个IP库，每个IP是否为代理IP，所处什么机房都有相应的记录，或者调用了IP位置查询接口，从而判断IP是否为代理IP、机房IP、个
第三个信息为攻击者第一次访问站点，从图中却到看到jsky的字样，竭思为一款Web漏洞扫描器，而根据我的经验来看，扫描器第一个请求不应该是访问一个txt文件而是应
第四、第五、第六个信息应该分别为访问了后台地址、对后台进行了爆破攻击、使用常见漏洞或CMS等通用漏洞对应用进行了攻击，除了后台访问成功之外，爆破攻击、应
第七个信息展示出了攻击者总共发起了79166次注入攻击，且对服务器已经造成了影响，但是从效果图中看来，此溯源并没有具体展示对哪台哪个应用攻击成功造成了影响，
对于图3中的效果，开始觉得结果丰富，意义深远，但是细看发现结果丰富大多来源于相关数据丰富。
综上所诉，此攻击溯源产品利用了两个优势得出了比常规分析日志方法中更有价值的结果
1.请求和响应数据完整，能进行更大维度的日志分析
2.安全关联库较多，能关联出更为丰富的信息
如下为产品中引用的关联库：

1. [redacted]IPV4[redacted]IP[redacted]IP[redacted]
2. [redacted]IP[redacted]IP[redacted]
3. [redacted]IP[redacted]IP[redacted]C&C[redacted]
4. [redacted]
5[redacted]IP[redacted]IP[redacted]
6. [redacted]

7. ■■
8. ■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■50■■■■■■■■■■■■■■■■■■■■

1. 永久记录黑客攻击的所有日志，为攻击取证溯源提供详细依据。

八、日志安全分析攻击溯源之路 [构想篇]

我也希望我在这一节能写出关于溯源的实践篇，然而事实是到目前为止，我也没有太好的办法来解决在传统日志分析中第三节中提到的问题，期间也做过一些尝试，得到的结
通过前几节，我们已经知道了我们分析日志的目的，攻击溯源的目的和其意义与价值
这里简短概括一下：
一、实时监控正在发生的安全事件、安全趋势
二、还原攻击者行为
1.从何时开始攻击
2.攻击所利用的工具、手法、漏洞
3.攻击是否成功，是否已经造成损失和危害
三、发现风险、捕获漏洞、修复漏洞、恶意行为取证

在传统日志分析过程中，想要实现以上效果，那么就不得不面对第三节中提到的问题，这里回顾一下：
1.POST数据不记录导致分析结果不准确
其实在服务器端，运维管理人员可自行配置记录POST数据，但是这里说的是默认不记录的情况，所以配置记录POST数据暂且不提
其实我觉得要从不完整的信息中，分析得到一个肯定的答案，我觉得这从逻辑上就不可行。但是我们可以折中实现，尽量向肯定的答案靠近，即使得到一个90%肯定的答案
在常规日志分析中，虽然POST数据不被记录，但是这些"不完整信息"依然能给我们我们提供线索
如通过响应大小、响应时间、前后请求关联、POST地址词义分析、状态码等等依然能为我们的分析提供依据，如某个请求在日志中的出现次数占访问总数30%以上，且响应

2.状态码不可信
对于那些自行设置响应状态的，明明404却302的，明明500却要200的(~~我能说这种我想拖出去打死么~~,) PS：其实设置自定义状态码是别人的正常需求
因为状态码不可信了，我们必须从其他方面入手来获取可信线索，虽然要付出点代价
我的思路是，对于不同的攻击行为，我们应该定义不同的响应规则，如攻击规则命中的为网站备份文件，那么应该判断请求大小必须超过1k-5k，如攻击者发起/wwwroot.ra

3.攻击者使用多个代理IP导致无法构成整个攻击路径
假设同一攻击者发起的每个请求都来自不同的IP，此时即使攻击规则命中了攻击者所有请求，也无法还原攻击者的攻击路径，此时我们只能另寻他法
虽然攻击者使用了多个IP，但是假设攻击者不足够心细，此时你可以通过攻击时间段、请求频率、客户端信息(Ua)、攻击手法、攻击工具(请求主体和请求来源和客户端信息中

4.无恶意webshell访问记录
常规分析中，我们通过找到后门文件，从而利用这一线索得知攻击者IP继而得知攻击者所有请求，但是如果我们并没有找到webshell，又该用什么作为分析的入口线索呢？
利用尽可能全面的攻击规则对日志进行匹配,通过IP分组聚合，提取发起过攻击请求的所有IP，再通过得到的IP反查所有请求，再配合其他方法检测提取出的所有请求中的可疑

5.编码避开关键字匹配
关于编码、加密问题，我也曾尝试过，但是实际最后发现除了URL编码以外，其他的编码是无法随意使用的，因为一个被加密或编码后的请求，服务器是无法正确接收和处理

6.APT分时段攻击
如果同一攻击者的攻击行为分别来源不同的时间，比如攻击者花一周时间进行"踩点"，然后他就停止了行为，过了一周后再继续利用所得信息进行攻击行为，此时因为行为链
其中每一个都可以理解一种行为，而每种行为都有相应的特征或者规则
比如主页链接一般在日志中占比较大，且通常路径为index.html、index.php、index.aspx,那么符合这两个规则则视为访问主页
而在探测注入行为中，一般会出现探测的payload，如时间注入会匹配以下规则：

```
.*(BENCHMARK\\(.*\\)).*
.*(WAITFOR.*DELAY).*
.*(SLEEP\\(.*\\).*
.*(THENDBMS_PIPE.RECEIVE_MESSAGE).*
```

Bool注入

```
.*and.*(>|=|<).*
   .*or.*(>|=|<).*
   .*xor.*(>|=|<).*
```

联合注入：

```
.*(order.*by).*
       .*(union.*select).*
       .*(union.*all.*select).*
       .*(union.*select.*from).*
```

显错注入:

```
.*('|"|\\)).*
       .*(extractvalue\\(.*\\)).*
       .*(floor\\(.*\\)).*
       .*(updatexml\\(.*\\)).*
```

利用注入则会体现出更完整，带有目的性的攻击请求，我们以同理制定规则即可，如查询当前数据库名、查询版本信息、查询数据库表名、列名则会出现database、version
扫描后台则会产生大量的404请求，且请求较为频繁，请求特征通常为/admin、/guanli、/login.php、/administrator
对于是否进入后台，我认为假如一个疑似后台访问的链接被频繁请求，且每次响应大小都不相同，我则认为这是已经进入了后台，但是也有可能是网站管理员正在后台进行操
关于上传webshell，这个比较难得到较准确的信息，因为我们没有POST数据，无法知道上传的内容是什么，但是我们可以通过反推法，先利用webshell访问特征进行匹配，
至于"通过webshell执行恶意操作"，可以简单定义为webshell地址被请求多次，且响应大小大多数都不相同
当我们对以上每种行为都建立对应的规则之后，然后按照攻击路径模型到日志中进行匹配，攻击路径模型可能有多个
这是一个相对常规的攻击路径：
访问主页>探测注入>利用注入>扫描后台>进入后台>上传webshell>通过webshell执行恶意操作
可能还会有
访问主页>爬虫特征>扫描敏感信息>扫描识别CMS特征>利用已知组件漏洞进行攻击>执行恶意代码>获取webshell>通过webshell执行恶意操作
扫描路径>扫描到后台>疑似进入后台>上传webshell>通过webshell执行恶意操作
..
当我们用多个类似这样的攻击路径模型对日志进行匹配时，可能在同一个模型中可能会命中多次相同的行为特征，此时我需要做一个排查工作，通过IP、客户端特征、攻击手
我们通过一整个攻击路径来定义攻击，从而即使攻击者分时段进行攻击，行为也会被列入到攻击路径中
通过这样方式，也许能实现自动化展示出攻击者的攻击路径，但是具体可行率、准确度还有待进一步实践后确认。

7.日志噪声数据
通常，除了攻击者恶意构造的攻击之外，日志中还包含大量的扫描器发出的请求，此类请求同样包含一些攻击特征，但是多半都为无效的攻击，那么我们如何从大量的扫描攻

九、日志安全分析之更好的选择 [大数据]

回到那个最基本的问题，如何从日志中区分正常请求和攻击请求？
可能做过安全的人都会想到：用关键字匹配呀
对，关键字匹配，因为这的确是简单直接可见的办法，用我们已知的安全知识，把每一种攻击手法定义出对应的攻击规则，然而对日志进行匹配，但Web技术更新速度飞快
其实从接触日志分析这个领域开始，我就想过一个问题？有没有一种算法，可以自动的计算哪些是正常的，哪些是不正常的呢？然而思索很久，也尝试过一些办法，比如尝试
后来又思索了一种办法，能不能对用户的网站产生的请求建立一个白名单，然后不在白名单内的请求皆为异常请求。这种做法效果倒是更好了一点，可是如何自动化建立白名
后来我发现其实我最初的想法其实是一个正确的思路，用统计的方法来区分正常和异常请求，只不过我在最开始实现的时候认为的是：某个URL被访问的次数越少，那么次请
更好的思路是：正常总是基本相似 异常却各有各的异常（来源：http://www.91ri.org/16614.html ）
文中关于此理论已经讲得很详细，这里简单描述一下实现方法：
搜集大量正常请求，为每个请求的所有参数的值定义正常模型
通过Waf或者攻击规则来剔除所有发起过攻击请求的IP，从而得到所有来自用户的正常请求，将每个正常请求构造出对应的正常模型，比如：
http://test.com/index.php?id=123
http://test.com/index.php?id=124
http://test.com/index.php?id=125
那么关于此请求的正常模型则为 [N,N,N],不匹配此模型的请求则为异常请求
当对日志中的请求建立完正常的模型，通过正常模型来匹配找出所有不符合模型的请求时，发现效果的确不错，漏报较少，不过实践中发现另一个问题，那便是数据的清洗，
关于此理论已经有人写出了Demo实现，地址：https://github.com/SparkSharly/Sharly

十、日志安全分析总结问答

1.日志分析有哪些用途？
感知可能正在发生的攻击，从而规避存在的安全风险
应急响应，还原攻击者的攻击路径，从而挽回已经造成的损失
分析安全趋势，从较大的角度观察攻击者更"关心"哪些系统
分析安全漏洞，发现已知或位置攻击方法，从日志中发现应用0day、Nday
..

2.有哪些方法可找出日志中的攻击行为？
攻击规则匹配，通过正则匹配日志中的攻击请求
统计方法，统计请求出现次数，次数少于同类请求平均次数则为异常请求
白名单模式，为正常请求建立白名单，不在名单范围内则为异常请求
HMM模型，类似于白名单，不同点在于可对正常请求自动化建立模型，从而通过正常模型找出不匹配者则为异常请求

3.日志分析有哪些商业和非商业工具/平台？
工具：
LogForensics 腾讯实验室
https://security.tencent.com/index.php/opensource/detail/15

北风飘然@金乌网络安全实验室
http://www.freebuf.com/sectool/126698.html

网络ID为piaox的安全从业人员：
http://www.freebuf.com/sectool/110644.html

网络ID：SecSky
http://www.freebuf.com/sectool/8982.html
网络ID：鬼魅羊羔
http://www.freebuf.com/articles/web/96675.html
平台（商业项目）：
Splunk >> 机器数据引擎

赛克蓝德 >> SeciLog
优特捷信息技术 >> 日志易
HanSight瀚思 >> 安全易
百泉众合数据科技 >>LogInsight
江南天安 >> 彩虹WEB攻击溯源平台
开源项目：
elk
https://www.elastic.co
scribe
https://github.com/facebook/scribe
chukwa
http://incubator.apache.org/chukwa/
kafka
http://sna-projects.com/kafka/
Flume
https://github.com/cloudera/flume/
4.有哪些方法适合分析攻击是否成功？
Kill Chain Model

十一、扩展阅读

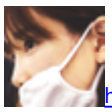
http://netsecurity.51cto.com/art/201506/478622.htm
http://www.freebuf.com/articles/web/86406.html
https://wenku.baidu.com/view/f41356138bd63186bdebbca8.html
http://www.freebuf.com/articles/web/96675.html
http://dongxicheng.org/search-engine/log-systems/
http://www.361way.com/scribe-chukwa-kafka-flume/4119.html
http://www.jianshu.com/p/942d1beb7fdd
https://xianzhi.aliyun.com/forum/attachment/big_size/WAF%E6%98%AF%E6%97%B6%E5%80%99%E8%B7%9F%E6%AD%A3%E5%88%99%E8%A1%A8%E8%
http://techshow.ctrip.com/archives/1042.html
http://www.ixueshu.com/document/b33cf4addda2a27e318947a18e7f9386.html
http://www.ixueshu.com/document/602ef355997f4aec.html
http://xueshu.baidu.com/s?wd=paperuri%3A%288b49643ad2a4ba7ea2d4cf46e366188d%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=h
□
□

十二、结束语

在安全领域中，防护为一个体系，感知风险和应急响应仅仅只算安全体系中的两个环节。而想要尽可能更好的实现这两个环节，单凭从日志中分析数据是远远不够的。
至于未来或许我们可以将日志分析和Waf、RASP、等其他安全产品进行联动，还可以将Web日志、系统日志、数据库日志等各种其他日志进行关联从而分析更准确、更具有
日志分析本质为数据分析，而数据驱动安全必定是未来的趋势。
关于日志分析还有很远的路要走，目前国内还并没有发现较为优秀的产品，日志数据中存在的价值还有待进一步挖掘。

上一篇：Mysql提权(CVE-2016-... 下一篇：Samba远程代码执行漏洞(CVE...

1. 37 条回复


hades 2017-06-14 07:14:02

有了新的见识 哈哈 看后续大伙如如何完善

0 回复Ta


master 2017-06-14 07:19:44

我草。年度好文啊。

0 回复Ta


wolf 2017-06-14 09:41:54

不错

0 回复Ta


御剑江湖 2017-06-14 09:53:47

不错不错，挺全的，现在对于日志分析难点还是在于如何自动化区分扫描式攻击和有效攻击，单纯从文本特征还是有限的，所以我觉得数据建模和文本特征相结合是一个

0 回复Ta


帅老头 2017-06-14 10:16:05

这个得 mark一下

0 回复Ta


hades 2017-06-14 10:27:04

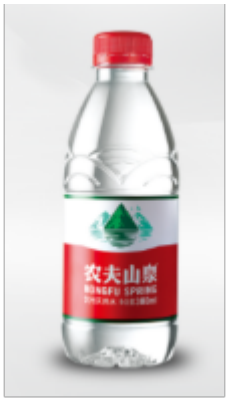
作者本意想结合机器学习来的

0 回复Ta


funk 2017-06-14 10:30:36

jeary 大表哥厉害。

0 回复Ta



anx1ang 2017-06-14 10:56:05

顶顶顶！~

0 回复Ta


evi1sly 2017-06-14 12:22:19

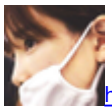
年度好文

0 回复Ta


笑看天下 2017-06-14 14:01:39

厉害了大牛

0 回复Ta


hades 2017-06-15 00:57:57

日志分析本质为数据分析，而数据驱动安全必定是未来的趋势。

我很认同这段话，后期会不会有专门的分析员出现？？

0 回复Ta

北风飘然 2017-06-15 02:22:24

之前也尝试做过日志分析
其实post请求如果是注入的话可以通过访问同一地址 请求次数与频率来看 (当然这是最初级的)
代理ip其实可以通过定位攻击时间段来判断是否是一个人所谓 ,但追溯就另说了
而无恶意webshell访问记录 个人认为攻击行为肯定会有的 只要有一条攻击行为就能大致定位到攻击的请求
而且还遇到恶心的地方比如安全狗拦截 iis日志里会显示状态码为200 ==
其实很尴尬的地方在于 判断到攻击行为了 怎么去辨别是否攻击行为生效这个问题 ==

以上是相对于日志较少的 个人观点 没遇到过大场面勿喷

0 回复Ta

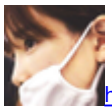
云卷云舒 2017-06-15 03:34:44

0 回复Ta

cover 2017-06-15 05:28:33

正好最近在做elk 日志分析，来得早不如来得巧

0 回复Ta

 hades 2017-06-15 06:24:30

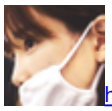
https://kibana.logstash.es/content/

0 回复Ta

 cover 2017-06-15 06:53:21

虽然已经搭好了， 还是要感谢老铁一波

0 回复Ta

 hades 2017-06-15 08:42:09

没毛病 走起

0 回复Ta

 文雨 2017-06-16 01:47:39

这个必须支持一下，写的确实不错，也做过类似的事情的，但是深度离这个很远

0 回复Ta


大亮_new 2017-06-20 05:54:28

日志的数据挖掘和大数据分析。

0 回复Ta


该隐 2017-06-22 15:14:49

正常总是基本相似 异常却各有各的异常，这句话一语中的~~~如何通过数据清理建立正常的模型，确实是值得思考的~

0 回复Ta


hades 2017-06-30 01:29:12

有什么好的想法可以探讨探讨ing

0 回复Ta

若水行 2017-07-11 07:18:00

学习了！

0 回复Ta

烤冷面加培根 2017-07-19 02:55:41

看到的太晚了，看到一些思路也是明白，还有这种操作？，楼主碰到的一些问题也曾想过办法，针对POST数据就是配置一个IDS，做一个筛子，大的POST数据包直接扔
对于第五个问题，匹配也许是日志分析中的最简单或许也是比较有效的解决办法
其他问题的话，还是宁可错杀也不放过的原则。
“至于未来或许我们可以将日志分析和Waf、RASP、等其他安全产品进行联动，还可以将Web日志、系统日志、数据库日志等各种其他日志进行关联从而分析更准确、更
数据库那一部分是一个另一种方法，本来想要集成到ELK的，但是自己实在是不会，溯源现在还没有涉及到，最后还是感谢楼主，有了这么一篇十分好的日志分析进阶文章

0 回复Ta

simeon 2017-07-19 08:02:52

牛逼文章！

0 回复Ta


bb好气 2017-07-24 07:32:13

我知道 我又 学习了 一波

0 回复Ta


我就是日志易 2017-07-27 11:06:38

您好，内容不错，很用心。不过内容的时效性有误差哦~

截图出现的日志易产品至少是一年半了，如果需要最新资料，欢迎致电联系日志易 contact@yottabyte.cn

或者私信撩起来

0 回复Ta


hades 2017-07-31 00:41:14

我去找找作者聊起来~

0 回复Ta


酸奶、 2017-08-06 02:12:29

写的不错。。。。

0 回复Ta


le0nis 2017-10-21 02:06:27

LZ 真的用心 ！！！6666，想知道 elk 更多的一些信息 kibana 最后是怎么演变成你图上面的那种效果的啊

0 回复Ta


le0nis 2017-10-23 08:34:28

楼主大表哥，第六章中的 将攻击规则应用于logstash的filter插件 具体是怎么实现的呢 求分享，并没有attackfilter 这个插件啊
和下面的过滤文件是怎么关联的呢

0 回复Ta


le0nis 2017-10-23 09:01:09

我明白了 这是自己开发的查件吧

0 回复Ta


gdyhw 2017-10-26 02:53:10

good

0 回复Ta


steven1881 2017-10-27 06:49:22

借鉴一下

0 回复Ta


鹰城广场 2017-10-28 08:54:45

厉害厉害 深度好文

0 回复Ta


jeary 2017-10-31 05:59:50

有相关数据就能直接在kibana里面建立视图。
你可以尝试搭建ELK然后导入常规的nginx日志，然后尝试建立饼图或者折线图来展现每天的访问量~
（然后发现，其实并不难）

0 回复Ta


le0nis 2018-01-16 16:45:05

大表哥，能把你匹配规则贡献一下不？？？？

1 回复Ta

 hades 2018-01-29 11:23:49

https://github.com/anbai-inc/AttackFilter 插件已经开源 @le0nis

0 回复Ta

登录 后跟帖

先知社区

现在登录

热门节点

技术文章

社区小黑板

目录

RSS 关于社区 友情链接 社区小黑板