安全小课堂第112期【安全日志分析系统建设】

京东安全应急响应中心 9月25日

日志分析在入侵检测中的应用越来越广泛。尤其在甲方安全的建设当中，由于种种原因，数据只能从日志当中去获取，这也给安全日志分析领域带来了一线生机，使其得以发展。在甲方安全工作当中，会使用分析日志的方式去发现日志当中存在的攻击行为或分析防火墙、安全设备、WAF、HIDS等产生的攻击日志，最终使得这些数据能够产生价值，帮助判读攻击行为及关联kill chain上下文信息。随着AI时代的到来，目前业内有一种说法叫AISEC，即通过AI（人工智能）及机器学习的方式去训练算法，帮助识别攻击行为，非常火热。

JSRC **安全小课堂第112期**，邀请到**Rozero**作为讲师就**安全日志分析系统建设技术**为大家进行分享。同时感谢各位小伙伴们的精彩讨论

安全日志分析系统的介绍？

京安小妹

Rozero：

安全日志分析系统在业内也叫SIEM或SOC，该系统的具体作用是通过收集生产环境中的日志。做分析后集中展示，搜索，告警，关联分析等功能，以实现对黑客攻击发生时或发生后的事件追溯及溯源需要或者业务安全监控需要。业内比较有名气的SIEM或SOC，有splunk或HP的arcsight，我只使用过其中的试用版本或者免费版本。在互联网企业中数据比较多的企业或者场景相对复杂的公司，都会考虑采用自己研发日志分析系统。成本可高可低。简而言之，如应用得当安全日志分析系统是一个可以全盘观察公司生产环境安全情况的工具。安全日志分析系统，在大的甲方或者爱研究的地方，多是自研。

讲师

安全日志分析系统数据是如何收集的？

京安小妹

Rozero：

数据收集采用的方式多种多样：

1.比如要收集网络数据netflow，需要用tap（Test Access Point）网络设备来采集网络流量，或镜像流量来采集网络访问请求

2.要采集主机的日志，可以通过syslog，rsyslog服务来采集搭建日志收集服务器，由client发送数据到日志收集服务器完成日志收集。

3.另外一种方式是采用Q（queen）来将数据写入到消息队列里，通过消费数据来完成日志的收集。

4.业务日志可以从数据库里读，类似分析用户的券领用，登录，注册等数据。

讲师

安全日志分析系统数据是如何处理的？

京安小妹

Rozero：

数据分析第一要考虑的是格式，如果是杂乱无章的数据是没办法梳理出头绪来分析的。第二需要能够理解业务，理解每行的日志对应的是什么意思。在对应的对日志的格式做切分，比如，在日志输出时即对日志格式做了定义以“\t”或“\r”来做切分，那么在输出的时候，通过“\t”或“\n”分割格式即可，对日志做解析。总之第一步，需要先能够理解数据，并且对日志做一定对预处理。

讲师

安全日志分析系统数据是如何分析的？

京安小妹

Rozero：

分析数据的方式也是五花八门，比较少的数据可以编写个shell、python脚本来实现。比如大的数据，现在也有很多实时计算的框架可以用于实现实时和离线（offline）分析。或者你可以用ELK（elasticsearch、logstash、kibana）来做相应的数据。又或者如果你有预算的话，也可以购买splunk，或者阿里云（SLS）服务及对应的odps服务来负责日志的分析。

讲师

安全日志分析系统数据是如何使用的？

京安小妹

Rozero :

1.数据的使用主要分为两种模式，1是通过把数据导入到hadoop hdfs里做离线分析，这个时候只需要熟悉一些常见的SQL语法即可通过SQL语法及恰当的规则表达式来实现对日志的检索分析来获取想要的结果。2.通过将想要描述的数据，通过计算机能够理解的语言编写成代码，做成实时计算任务，让程序自动去读取Q(queen)或者日志服务(elasticsearch) 来做分析。这里比较依赖专家经验。

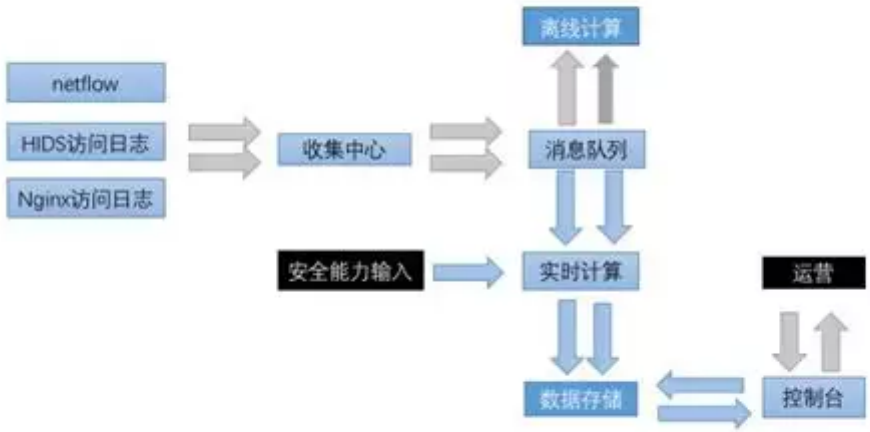
讲师

安全日志分析系统数据架构是什么样的？

京安小妹

Rozero：

通用的安全日志分析系统的架构相对来说比较简单。无非是日志收集、消费、存储，使用的场景。如下图是一个相对来说比较简单易懂的流程图。



如果没有人力物力，其实使用ELK（elasticsearch、logstash、kibana）这种架构来组成，相对来说简单易懂容易维护，缺点也很明显，比较ELK是免费的解决方案，你要自己踩很多坑。

讲师

互动问答环节：

1. 日志分析能适用到哪些场景呢？

讲师:

日志分析适用的场景是比较多的，举个例子，日志是一切的前提。不说在安全领域，日志可以用来debug，可以用来定位问题。而在安全领域，dns log、newflow、ids、hids这些日志都能够分析出很多有用的数据

2. 如何从日志分析中发现powershell的攻击行为？

讲师:

一般都是用hids来对powershell发起的行为做检测。或者检索eventlog。

3. netflow、HIDS、nginx日志、系统日志，有可能设备不同，存储位置不同，怎么统一收集呢？

讲师:

一般要分情况来收集存储。存储可以都存储在kafka，或者类似的Q里。但收集本身就是个体力活,很难集中,即使类似商业化的产品，也很难做到集中收集。做的比较好的是，你

配置个syslog把数据放过去。

4. 系统日志，如果通过rsyslog收集，如果要同步到多个rsyslog server上，该怎么设置呢？？

讲师:

安全一般也只关心secure的登录日志。而如果你的日志是自己定义的，那么可以在rsyslog里写个配置，来将该数据的内容发送过去就好。数据发送逻辑：rsyslogd=>rsyslog server *.* 是全部日志，你可以自己过滤的，syslog有level等级，具体可以看下syslog的说明。

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