ElasticSearch课题结题报告

目录

[摘要： 1](#_Toc503521309)

[课题报告： 2](#_Toc503521310)

[一、工时报告 2](#_Toc503521311)

[二、工作内容报告 2](#_Toc503521312)

[1、涉及ES的修改 2](#_Toc503521313)

[2、未涉及ES的修改 3](#_Toc503521314)

[三、对应的镜像改动报告 3](#_Toc503521315)

[四、各模块功能描述报告 3](#_Toc503521316)

[五、MYSQL数据存储报告 4](#_Toc503521317)

[六、ES数据存储结构报告 5](#_Toc503521318)

[七、ES数据来源报告 6](#_Toc503521319)

[八、代码结构报告 7](#_Toc503521320)

[总结 7](#_Toc503521321)

# 摘要：

数据存储方案替换（由MYSQL切换为ElasticSearch）的课题开发工作已全部完成，现在我们的系统并存了两种数据存储服务，MYSQL与ES，绝大部分数据的存储都已移植到ES，但流量这一部分的数据因需求的原因导致难以移植到ES，尤其在10G的生产环境下，ES的写入性能难以满足我们的需求，所以这一部分仍采用MySQL作为数据存储引擎，告警的数据以及系统信息、配置的数据都用ES存储，从目前的情况来看，ES在查询上的表现良好，后续会对其进行稳定性测试，以及长时间运行后系统的负载情况等。

# 课题报告：

## 一、工时报告

预期工时：38人天

实际消耗：28人天

## 二、工作内容报告

### 1、涉及ES的修改

1.1、digger引擎

可疑域名、IP、URL的数据入库；

端口流量小时统计数据入库；

端口流量异常报警数据入库；

白名单域名对应的白名单ip入库；

ddos攻击的报警数据入库；

地域流量的上行数据入库；

攻击成功判定的信息入库；

1.2、portrait

外部攻击告警数据入库；

地域流量异常报警数据入库；

自动检测到的资产ip入库；

攻击成功状态数据来源修改；

策略集以及告警白名单数据来源修改；

1.3、alerting

告警数据来源修改；

邮件配置来源修改；

1.4、web

php整体修改

1.5、Python

会话异常判定，告警数据入库；

1.6、系统数据库初始化脚本修改

建立index以及type的mapping；

初始化文档数据，如规则信息，攻击类型信息，

1.7、审计日志修改

1.8、系统镜像制作

### 2、未涉及ES的修改

2.1、surveyor

surveyor移除数据库操作；

2.2、squirrel

修改ring-buffer，发送数据到portrait；

2.3、portrait

增加ring-buffer，接收数据并处理；

2.4、minerva

添加地域信息数据并入库，生成页面中流量查询的数据源

2.5、es-c-api

新增es的c接口，封装es的操作以及对返回结果的处理接口

## 三、对应的镜像改动报告

1、JDK1.8

2、ElasticSearch5.6.3并随系统启动

3、php由之前的php5.5升级到php5.6

4、Apache启用反向代理，代理9200端口，ElasticSearch的默认端口修改为6201

5、增加es-php接口插件

6、增加php-curl插件

7、mysql版本升级为5.6

8、增加curl以及Apache2-utils

## 四、各模块功能描述报告

以下主要说明各模块的主要功能，并不代表所有功能。

1、mgmt管理

1.1、启动与管理系统其它子进程（模块）；

1.2、与其他模块的心跳保持；

1.3、提供页面与底层通信的接口。

2、surveyor入侵检测引擎

2.1、检测系统是否受到了外部攻击，若检测到攻击则记录攻击数据到日志文件中；

2.2、将捕获到的流量信息发往digger以及minerva。

3、squirrel日志解析

3.1、解析surveyor引擎产生的告警日志；

3.2、将解析的结果发往其他portrait进程进行入库处理。

4、portrait数据入库

4.1、处理squirrel发送过来的数据，并在一定的过滤（策略过滤，白名单过滤） 之后，将告警数据入库；

4.2、检视统计到的地域流量的数据是否异常，若异常则将异常数据入库；

4.3、资产的自动发现。

5、digger异常流量检测引擎

5.1、检测http协议与dns协议的数据是否存在异常，如可疑ip、域名、url的检测， dns行为异常的检测，并将异常数据入库；

5.2、白名单ip的生成，会话异常判定需要根据该白名单进行过滤，降低误报率；

5.3、攻击成功与否的判定；

5.4、端口流量的小时数据统计以及端口异常告警）；

5.5、地域上行流量的统计。

6、alerting邮件告警

6.1、对攻击成功的事件向页面配置的收件人发送告警邮件。

7、minerva流量统计引擎

7.1、对资产ip，常见端口（包含配置端口），协议、地域信息进行流量统计；

7.2、对会话数据进行统计；

8、sf\_aly流量分析

8.1、对统计的会话数据进行分析，并将可疑的会话数据入库；

8.2、对地域信息（国家省市经纬度）入库。

## 五、MYSQL数据存储报告

由于es难以满足对流量数据的组合查询（按时间、资产、协议、端口、地域），以及会话数据的管理，所以这一部分的需求还是按照之前mysql的处理方式完成。

目前用到mysql的部分：

1、流量查询的数据源；

2、系统信息的数据源；

3、流量分布统计的数据源；

为方便查看，此处留白

## 六、ES数据存储结构报告

主要是对每一个index下有哪些type，每个type是干什么用的做一个全局说明。

|  |  |  |
| --- | --- | --- |
| Index | Type | 备注 |
| events | yyyymm | type是年月，如201710，包括外部攻击告警（surveyor触发规则）以及内部资产对可疑IP、域名、url访问产生的告警数据 |
| behavior | yyyymm | type是年月，如201710，存储行为异常的数据，包括端口流量异常、地域流量异常、ddos攻击、会话异常的数据 |
| utils | keyword\_matched | 攻击成功判定匹配到的数据 |
| property\_info | 资产信息数据 |
| ip\_domain\_whitelist | Ip域名白名单 |
| email\_alerts | 邮件告警配置 |
| policy\_set | 策略集 |
| alerts\_ip\_whitelist | Ip告警白名单 |
| attack\_type | 攻击类型翻译数据 |
| rule\_info | 规则信息数据 |
| user | web用户 |
| role | Web角色 |
| system\_utils | 系统信息 |
| port\_config | 端口配置 |
| spooler\_accumid | accum\_ids\_1 | 存储cid |
| traffic | baseline | 资产不同时段的流量均值与方差 |
| uplink\_traffic | 地域上行流量小时数据 |
| port\_traffic | 端口流量小时数据 |
| geo\_state | geo\_info\_map | 国家省市信息 |
| audit\_log | yyyymm | type是年月，如201710，页面操作审计信息 |
| sys\_audit\_log | db\_audit\_log | 数据库更新或者删除审计信息 |

为方便查看，此处留白

## 七、ES数据来源报告

主要是对每一个功能模块用到的哪些index、type做一个全局的说明。

|  |  |  |
| --- | --- | --- |
| Module | Index/Type | 备注 |
| Portrait | Events/yyyymm | Surveyor产生的告警 |
| Traffic/baseline | 地域流量不同时间段的均值方差 |
| Behavior/yyyymm | 地域流量异常告警 |
| Utils/property\_info | 资产自动发现 |
| digger | Events/yyyymm | 可疑域名/ip/url产生的告警 |
| Utils/ip\_domain\_whitelist | Digger生成的Ip域名白名单 |
| Utils/matched\_keyword | 攻击成功判定 |
| Traffic/port\_traffic | 端口小时流量统计（用于异常判定） |
| Behavior/yyyymm | Ddos攻击判定 |
| Traffic/uplink\_traffic | 地域小时流量（用于异常判定） |
| Squirrel | spooler\_accumid/ accum\_ids\_1 | Id在多台设备上部署可能会改变，目前id为1，如果以后要部署多台，注意es\_init.sh建mapping的写法 |
| Sf\_aly（Python） | Behavior/yyyymm | 异常会话产生的告警 |
|  | Geo\_state/ geo\_info\_map |  |
| Db\_init | Utils/Attack\_type | 安装软件初始化数据，页面不能配置 |
| Utils/rule\_info |
| Utils/user | 安装软件初始化数据，页面可以配置,与Web模块共享 |
| Utils/role |
| Utils/system\_utils |
| Utils/port\_config |
| Web | Utils/property\_info | 页面配置 |
| Utils/email\_alerts |
| Utils/policy\_set |
| Utils/alerts\_ip\_whitelist |
| Audit\_log/yyyymm | 页面访问审计 |
| apache | Sys\_audit\_log/db\_audit\_log | 数据库更新删除审计 |

为方便查看，此处留白

## 八、代码结构报告

代码主要分为七个部分。

|  |  |  |
| --- | --- | --- |
| 分支名 | 代码目录名 | 备注 |
| PUBLIC\_LIB | curl | Curl库 |
| Es-c-api | Es的C接口库 |
| Json-c | Jsonc库 |
| heartbeat | 心跳包库 |
| libmaxminddb-1.0 | 地域信息库 |
| Libdumbnet，libpcap，libpcre | Surveyor依赖的库 |
| SUR\_GENERAL | daq | Surveyor与抓包引擎的中间件 |
| pfring | Pfring抓包引擎 |
| minerva | 流量统计引擎 |
| squirrel | 日志解析 |
| Sur\_netset | 内网配置 |
| Sur\_user\_rules | 用户自定义规则 |
| surveyor | 入侵检测引擎 |
| Sys\_info | 获取系统信息 |
| Upgrade | 系统升级 |
| PSI\_INSTALL | 安装脚本，流量分析脚本 |
| SURVEYOR\_ES | Alerting | 邮件告警 |
| Digger | 异常流量检测引擎 |
| Management | 系统功能管理模块 |
| Portrait | 告警入库模块 |
| Recorder | 流量记录，未用 |
| Statistics | 协议、端口流量统计，未用 |
| WEB | Web\_es | Web代码 |
| DB\_INIT | db\_init\_es | 数据初始化代码 |
| rules\_v102 | rules\_v102 | surveyor规则 |
| Usi\_midware | Usi\_midware | Dpdk |

# 总结

首先感谢大家的支持，在大家的共同努力下，我们完成了这一课题的开发工作，在这一段开发时间里，遇到一些困难，但我们都一起商量讨论，拿出解决方案并在工作中顺利实施，确保了系统的交付，后续我们还会有一部分工作在解bug以及优化上，还请各位继续对该课题的代码以及功能进行维护，谢谢。