# （I）变化资产

## 一、设计原理

如果某ip短时间内同时发生多种类型事件（端口异常、异常流量、应用行为异常），则认为该ip为变化资产。

应用行为异常：该ip的操作系统、mac地址发生变化（无相关数据，暂时没考虑）

## 二、数据源

### （1）训练数据

#### 训练样本生成/采集方式

无

#### 数据源地址

无

#### 数据源格式

无

### （2）检测数据

#### 检测数据生成/采集方式

提取1d内ES的event\_yyyyMMdd中的event\_rule\_name ={端口异常、异常流量}

#### 数据源地址

es的事件表（event\_yyyyMMdd）

es.nodes = 10.130.10.18

es.port =9200

#### 数据源格式

表2 event\_yyyyMMdd

|  |  |  |
| --- | --- | --- |
| ID | event\_rule\_name | destip |
| 唯一标识 | 端口异常/异常流量 | 发生的网元ip地址 |

## 三、特征处理

读取当天es的事件表中event\_rule\_name、destip字段，筛选。

## 四、算法模型

### 算法名称

基线

### 算法详情

基于同一个ip短时间内不可能同时触发多个事件，故此考虑1d检测一次是否有变化资产：

a. 读取es的事件表（event\_20180719） event\_yyyyMMdd/event\_yyyyMMdd（event\_rule\_name（端口异常(destip)、异常流量(destip)））1天内上报的IP;

b.利用应用行为基线模型检测的应用行为异常ip，结合上一步骤a读取es事件表的相关ip，进行聚合操作;

c. 一天检测1次，聚合ip当天所对应的事件

|  |  |
| --- | --- |
| ip | events |
| IP1 | {异常流量、端口异常} |

d．设定阈值a =1 ，当ip.events >a, 则该ip为变化资产。

## 五、训练过程详情步骤

### （1）训练周期

### （2）训练步骤

## 六、检测过程详情步骤

### （1）检测周期

每天凌晨00:30

### （2）检测步骤

Step1：

读取前1d的ES事件表中{端口异常、异常流量}，聚合各个ip发生的事件，如果该ip发生的事件种类超过阈值（默认为1），则认为该ip为变化资产

## 七、输出结果

### （1）模型保存位置

无

### （2）检测结果

#### 保存方式 postgresql

#### 保存位置

postgre.address =jdbc:postgresql://10.130.10.22:5432/BDSSA1

如有变动，请自行修改

#### 结果格式

T\_SIEM\_DEV\_FIND

|  |  |  |  |
| --- | --- | --- | --- |
| recordid | IP | STORAGETIME | FLG |
| 唯一标识 |  | 入库时间 | "flg" IS '4变化资产' |

#### 结果示例

略

## 模型评估

略

# （II）闲置资产

## 一、设计原理

通过一段时间内判断低于某个流量阀值的资产，认为该资产是闲置资产。（cpu和内存利用率，待考虑）

## 二、数据源

### （1）训练数据

#### 训练样本生成/采集方式

通过将T\_HISTORY\_MODELRULES的当天（TYPE =异常流量 && IDENTITYRESULT= ip\_data && STANDBY03=tcp）数据，提取相关字段，利用算法处理后生成新的记录保存至T\_HISTORY\_MODELRULES的当天（TYPE =待检测闲置资产 && IDENTITYRESULT= history\_asset）。

**数据源地址**  
postgre.address =jdbc:postgresql://10.130.10.9:5432/BDSSA1

ip如有更换，请自行修改

#### 数据源格式

数据库表T\_HISTORY\_MODELRULES

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ID | TYPE | TYPERESULT | IDENTITYRESULT | STANDBY02 | STANDBY03 | baselineresult |
| 唯一标识 | 异常流量 | 时间（精确到5min，格式为yyyymmddhhmm） | ip\_data | 当前5min内出现的所有网元ip | tcp | 当前5min内相关网元ip对应的字节 |

### （2）检测数据

#### 检测数据生成/采集方式

读取PG上T\_HISTORY\_MODELRULES的当天（TYPE=待检测闲置资产 && IDENTITYRESULT= history\_asset）数据。

#### 数据源地址

postgre.address =jdbc:postgresql://10.130.10.9:5432/BDSSA1  
ip如有更换，请自行修改

#### 数据源格式

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | IP | TYPE | TYPERESULT | IDENTITYRESULT | STANDBY01 |
| 唯一标识 | 当天出现的资产ip | 待检测资产 | 访问host的日期（格式为yyyymmdd） | history\_asset | 入库时间（long类型然后存string） |

## 三、特征处理

1. 记录时间：提取出时间的分钟时段，以5分钟之内做统计，假设某个时间是2017-07-19 17:42:32，则提取的时段为2017-07-19 17:40:00。

由于记录数据的缺点，会造成同一次传输被分开记录，因此将5分钟之内的所有数据统计出来。

1. 字节数：针对内网ip段（172.16.21~、172.16.10~）进行过滤，初期数据如下表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| RecordTime | SrcIP | DstIP | Ups | Downs |
| 2017-07-19 17:40:00 | 内网ip（172.16.21.248） | - | 62 | 0 |
| 2017-07-19 17:40:00 | - | 内网ip（172.16.21.248） | 0 | 100 |

参考上表格，如果SrcIP是内网ip，只考虑Ups；如果DstIP是内网ip，只考虑Downs。则该时间段（5分钟内）可以将其聚合为

|  |  |  |
| --- | --- | --- |
| IP（内网ip） | Bytes | Protocol |
| 172.16.21.248 | 162（62+100） | - |

可以参考（流量异常检测模型->特征处理）以tcp协议考虑。

根据之前的流量异常模型5min存入Postgresql的表t\_history\_modelrules，筛选出当天（type=’异常流量’and standby03=’tcp’，理论上有288个5min的值）的数据standby02代表ips（这5min的网元ip用“#”合并，“172.16.12.1#172.16.12.2”），baselineresult代表各个ips对应的字节值(示例：“530.00#540.00#550.00”）。经过数据预处理得到如下表

|  |  |
| --- | --- |
| ip | bytes\_value（1天288个值） |
| IP1 | [v1, v2, v3, …, v287, v288] |
| IP2 | [v1, v2, v3, …, v287, v288] |

设定流量阈值b（1kb默认1024），如果ip.bytes\_value都小于b，则认为该ip为待检测闲置资产，更新入库至历史表t\_history\_modelrules（TYPE =待检测闲置资产 && IDENTITYRESULT= history\_asset）数据。

## 四、算法模型

### 算法名称

规则

### 算法详情

当前经过特征处理后，刚得到的待检测闲置资产ip与t\_history\_modelrules（type=‘待检测闲置资产 && identityresult=‘history\_asset’）的ip对比，如果同个ip的（TYPERESULT字段）发生时间差超过阈值（默认3day），则为该ip为闲置资产，入库T\_SIEM\_DEV\_FIND；最后将刚得到待检测闲置资产ip更新入库t\_history\_modelrules。

## 五、训练过程详情步骤

### （1）训练周期

每天凌晨00：30运行

### （2）训练步骤

Step1：

读取T\_HISTORY\_MODELRULES的当天（TYPE =异常流量 && IDENTITYRESULT= ip\_data && STANDBY03= tcp）数据；

Step2：聚合各个ip当天的值，保存至数组ip. bytes\_value=[v1,v2,v3,…,v288]；

Step3：ip.bytes\_value都小于设定阈值（1kb=1024），则认为该ip为待检测闲置资产。

## 六、检测过程详情步骤

### （1）检测周期

每天凌晨00:30开始运行

### （2）检测步骤

Step1：将训练步骤中检测出来的待检测闲置资产ip与t\_history\_modelrules（type=‘待检测闲置资产 && identityresult=‘history\_asset’）的ip对比，如果同个ip的（TYPERESULT字段）发生时间差超过阈值（默认3day），则为该ip为闲置资产。

## 七、输出结果

### （1）模型保存位置

无

### （2）检测结果

#### 保存方式 postgresql

#### 保存位置

postgre.address =jdbc:postgresql://10.130.10.22:5432/BDSSA1

如有变动，请自行修改

#### 结果格式

T\_SIEM\_DEV\_FIND

|  |  |  |  |
| --- | --- | --- | --- |
| recordid | IP | STORAGETIME | FLG |
| 唯一标识 |  | 入库时间 | "flg" IS ‘5闲置资产' |

#### 结果示例

略

## 模型评估

略

# （III）退网资产

## 一、设计原理

通过对比资产表，该ip过了一段时间（1个月或半年）后才重新出现，则认为该资产是退网资产。

## 二、数据源

### （1）训练数据

#### 训练样本生成/采集方式

通过将T\_HISTORY\_MODELRULES的当天（TYPE =异常流量 && IDENTITYRESULT= ip\_data && STANDBY03=tcp）数据，提取IP字段，整合成新的记录更新入库至T\_HISTORY\_MODELRULES的当天（TYPE =待检测退网资产 && IDENTITYRESULT= history\_asset）。

**数据源地址**  
postgre.address =jdbc:postgresql://10.130.10.9:5432/BDSSA1

ip如有更换，请自行修改

#### 数据源格式

数据库表T\_HISTORY\_MODELRULES

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | TYPE | TYPERESULT | IDENTITYRESULT | STANDBY02 | STANDBY03 |
| 唯一标识 | 异常流量 | 时间（精确到5min，格式为yyyymmddhhmm） | ip\_data | 当前5min内出现的所有网元ip | tcp |

### （2）检测数据

#### 检测数据生成/采集方式

读取PG上T\_HISTORY\_MODELRULES的当天（TYPE =待检测退网资产 && IDENTITYRESULT= history\_asset）数据。一天出现多少个网元ip，就有多少条记录。

#### 数据源地址

postgre.address =jdbc:postgresql://10.130.10.9:5432/BDSSA1  
ip如有更换，请自行修改

#### 数据源格式

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | IP | TYPE | TYPERESULT | IDENTITYRESULT | STANDBY01 |
| 唯一标识 | 当天出现的资产ip | 待检测退网资产 | 访问host的日期（格式为yyyymmdd） | history\_asset | 入库时间（long类型然后存string） |

## 三、特征处理

建立历史资产：根据之前的流量异常模型5min存入Postgresql的表t\_history\_modelrules，筛选出当天（type=’异常流量’and standby03=’tcp’，理论上有288个5min的值）的数据，standby02代表ips（这5min的网元ip用“#”合并，示例:“172.16.12.1#172.16.12.2”）

将一天的standby02重新聚合去重，保存至表t\_history\_modelrules（ip表示当天出现的资产ip，type=‘待检测退网资产’， typeresult表示发生时间（格式为yyyymmdd）， identityresult=‘history\_asset’,standby01表示入库时间）

## 四、算法模型

### 算法名称

规则

### 算法详情

当前经过特征处理后，刚得到待检测退网资产ip与t\_history\_modelrules（type=‘待检测退网资产 && identityresult=‘history\_asset’）的ip对比，如果同个ip的（typeresult字段）发生时间差超过阈值（默认30day），则为该ip为退网资产，入库T\_SIEM\_DEV\_FIND；最后将刚得到待检测退网资产ip更新入库。

## 五、训练过程详情步骤

### （1）训练周期

每天凌晨00：30运行

### （2）训练步骤

提取昨天的T\_HISTORY\_MODELRULE（type=’异常流量’and standby03=’tcp’，理论上有288个5min的值），经过特征处理，更新入库至T\_HISTORY\_MODELRULES的当天（TYPE =待检测退网资产 && IDENTITYRESULT= history\_asset）

## 六、检测过程详情步骤

### （1）检测周期 与训练过程同时运行

### （2）检测步骤

Step1：

读取T\_HISTORY\_MODELRULES（TYPE =待检测退网资产 && IDENTITYRESULT= history\_asset）数据；

Step2：具体请参考算法详情

## 七、输出结果

### （1）模型保存位置

无

### （2）检测结果

#### 保存方式 postgresql

#### 保存位置

postgre.address =jdbc:postgresql://10.130.10.22:5432/BDSSA1

如有变动，请自行修改

#### 结果格式

T\_SIEM\_DEV\_FIND

|  |  |  |  |
| --- | --- | --- | --- |
| recordid | IP | STORAGETIME | FLG |
| 唯一标识 |  | 入库时间 | "flg" IS ‘3闲置资产' |

#### 结果示例

略

## 模型评估

略

# 总结：

1. 变化资产的数据来源为ES的事件表event\_yyyyMMdd中的event\_rule\_name ={端口异常、异常流量}，如果某ip当天同时发生{端口异常、异常流量}，则认为该ip为变化资产

2. 闲置资产的数据来源为T\_HISTORY\_MODELRULES（type=’异常流量’and standby03=’tcp’），根据相关算法处理成新的数据保存至T\_HISTORY\_MODELRULES（TYPE =待检测闲置资产 && IDENTITYRESULT= history\_asset）。然后根据这次发生的时间与之前发生的时间作比较，上报最终的闲置资产ip。

3. 退网资产的数据来源为T\_HISTORY\_MODELRULES（type=’异常流量’and standby03=’tcp’），根据相关算法处理成新的数据保存至T\_HISTORY\_MODELRULES（TYPE =待检测退网资产 && IDENTITYRESULT= history\_asset）。然后根据这次发生的时间与之前发生的时间作比较，上报最终的退网资产ip。

4. T\_HISTORY\_MODELRULES（type=’异常流量’and standby03=’tcp’）的数据是“流量异常模型”的特征处理方法，具体请参考文档；必须配合先运行<http://10.130.10.38:9999/svn/bdtsgz/branches/BDSSA/DataMining/Traffic_rebuild/src/main/scala/jobFiveMinute/subJob/netflowProcess/>，才能进行检测。

5. T\_HISTORY\_MODELRULES的历史数据需要定时删除，可以通过standby01字段（入库时间，long类型然后存string）进行筛选删除（好像还没有写过脚本，需另行开发，删除周期的话看情况咯，2个月删除一次呗）。

6．变化资产、闲置资产、退网资产最终的事件入库至T\_SIEM\_DEV\_FIND。