

机器学习&攻击检测

◆
携程 / 岳良



2018 携程安全沙龙

目录

ML&恶意域名检测

1

ML&web攻击检测

2

3

Question



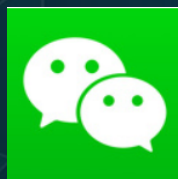
ML&恶意域名——检测需求的诞生

那一年应急带来的痛苦回忆

xshell , wannacry

依靠传统杀毒软件？

NO！还不如靠朋友圈



怎么解决

买威胁情报服务？

自研产品，猜测存在隐患的随机域名？



看下恶意域名的特点

yhxpqrhnhnwmvdcwj.eu
bxedpvqwnbrqilaawj.com

提炼特点

不可读
比较长

有的域名甚至都没注册

怎么检测

为什么人眼一看就知道哪些域名可疑



ML&恶意域名——DGA算法

The screenshot shows the GitHub interface for the repository 'banjori'. The top navigation bar includes 'Pull requests', 'Issues', 'Marketplace', and 'Explore'. The left sidebar contains navigation links for 'Repositories' (1), 'Code' (196), 'Commits', 'Issues' (7), 'Marketplace', 'Topics', 'Wikis', and 'Users' (4). Below the sidebar is a 'Languages' section with the following data:

Language	Count
Python	55
CSV	36
Markdown	18
Text	18
JSON	11
Jupyter Notebook	10
Shell	6
PHP	4
RMarkdown	4
YAML	4

The main content area displays '196 code results'. It features three search results for the file 'IOC_20170904.txt':

- ChrisLinn/greyhame-2017 – IOC_20170904.txt**
Showing the top four matches Last indexed on 12 Jul
- hao-alien/greyhame-2017 – IOC_20170904.txt**
Showing the top four matches Last indexed on 14 Jul
- quydx/dnschecklist – banjori_dga.csv**
Showing the top two matches Last indexed on 12 Jul

The search results for the first two repositories show a list of domains used by 'banjori':

- 1 aakamen.com,Domain used by banjori
- 2 aaskmen.com,Domain used by banjori
- 3 aifamen.com,Domain used by banjori
- 4 aigemen.com,Domain used by banjori
- 5 asismen.com,Domain used by banjori
- 6 aswomen.com,Domain used by banjori

The search results for the third repository show a list of domain patterns:

- 1 "andersensinaix.com","0","banjori_dga_andersensinaix.com_0x3c03"
- 2 "xjsrnsensinaix.com","1","banjori_dga_andersensinaix.com_0x3c03"
- 3 "hlrfrsensinaix.com","2","banjori_dga_andersensinaix.com_0x3c03"



ML&恶意域名——检测产品设计

字典匹配

携程内部的专有域名
alexa top 100万的域名
收集的100万黑域名

机器学习
检测

TFIDF提取特征训练模型

外部威胁
情报API
调用

调用某国外厂商接口做机器学习
结果的二次验证



ML&恶意域名——检测产品设计



ML&恶意域名检测——ML步骤

步骤

定义问题

数据收集, 清洗

特征工程

训练模型

评估模型效果, 改进

上线调用模型

随机域名识别的二分类问题

外部各100万黑白样本

tfidf提取

xgboost
算法

交叉验证,
线上测试

线上RUN

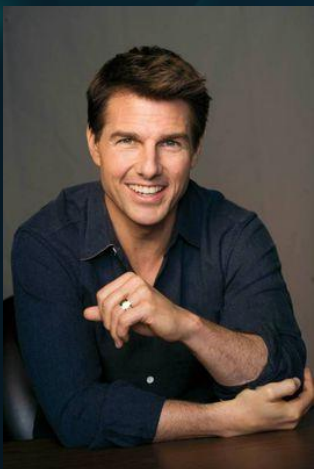


ML&恶意域名检测——总结人脑的工作流程

问题？小学生A今年7岁。有一天拿到了一篇不带标题的微信文章，请问其是如何分辨这篇文章的主题是“世界杯”还是“娱乐” **二分类问题**

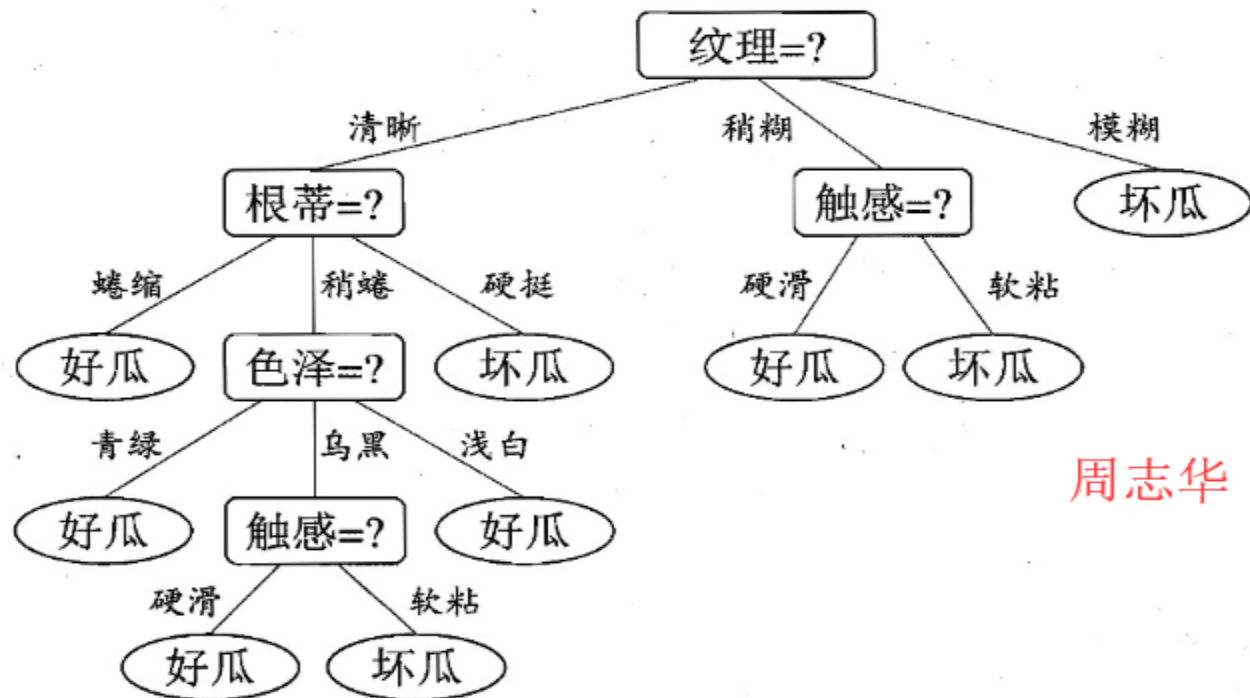
首先，如果小学生A家里没网没电视没报纸，那他肯定分辨不出来 **没有样本**

小学生A家里新装了宽带，但是A骨骼惊奇，将网上的世界杯和娱乐新闻全部转成了汉语拼音，然后得到结论，a ba fa de几个拼音出现多的就是世界杯 **特征提取有问题**



ML&恶意域名检测——总结人脑的工作流程

经过老师的指点，A重新改了特征（分词取汉语关键词出现频率），且大脑总结了一套算法，先看是否有一次“足球”，如果有再看有没有超过2次“梅西”，如果有就基本95%确定是世界杯新闻了 **训练并使用了“决策树”机器学习算法**



周志华《机器学习》决策树



国际足联/裁判委员会/主席/科里纳/在/总结/48/场/小组赛/时/表示，共用/VAR/查看了/335/次/犯规，每场比赛/接近/7/次/。这些/犯规/中，在/没有/VAR/的情况下，95%/的/判罚/是/正确/的，但/VAR/更正/了14/次/判罚，让/准确率/提升/到/99.3%。

在/今年/金马奖/官宣/前,巩俐/就/已经/先后/出任/过/第/50/届/戛纳/国际/电影节/主竞赛/单元/评委/、第/50届/柏林/国际/电影节/主席/、第/59/届/威尼斯/国际/电影节/主席。



ML&恶意域名检测——特征提取Tfidf

ctrip.com, ngram=2 ,按字符抽取

得到[ct, tr, ri, ip]

计算ct的词频TF

TF=ct在[ct, tr, ri, ip]出现的次数1/数组长度4=0.25

计算ct的逆文档词频IDF

IDF=log(训练语句总数/(含ct的语句个数+1))=log(10000/11)

TFIDF = TF * IDF



ML&恶意域名检测——模型效果

决策树

Precision: 0.933

Recall: 0.900

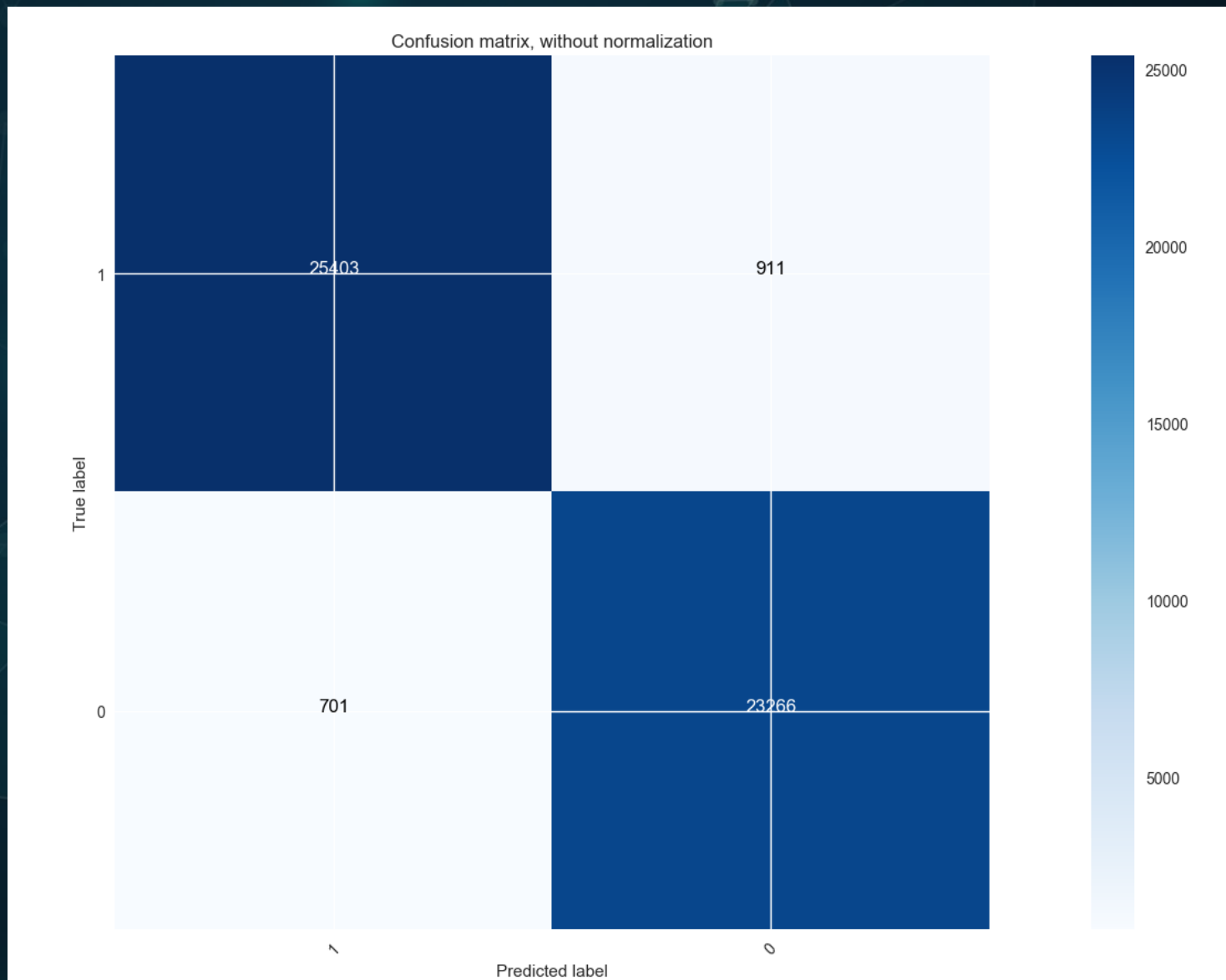
F1: 0.916

xgboost

Precision: 0.973

Recall: 0.965

F1: 0.969



说一下为什么3.5%的漏报率可以接受

sentqq.com


sexoqq.com

sexyos.net

goodfred431.com



ML&恶意域名检测——实际效果



时间: 2018/04/07 - 2018/09/01 域名: f582e064be47 服务器类

类别: 全部 状态: 全部

编号	时间	域名	次数	类别	状态	备注	操作
556967	2018/04/26 00:14:22	f582e064be47	90	机器学习	确认	N/A	<input type="button" value="标记"/> <input type="button" value="备注"/> <input type="button" value="推送"/>
554110	2018/04/25 09:02:47	f582e064be47	79	机器学习	确认	N/A	<input type="button" value="标记"/> <input type="button" value="备注"/> <input type="button" value="推送"/>
553646	2018/04/24 18:33:52	f582e064be47	6	机器学习	待定	N/A	<input type="button" value="标记"/> <input type="button" value="备注"/> <input type="button" value="推送"/>
551548	2018/04/23 16:32:29	f582e064be47	3	机器学习	待定	N/A	<input type="button" value="标记"/> <input type="button" value="备注"/> <input type="button" value="推送"/>
565017	2018/04/29 11:37:09	f582e064be47	1	样本数据	确认	http://camaradeideias.com.br/...	<input type="button" value="标记"/> <input type="button" value="备注"/> <input type="button" value="推送"/>
569527	2018/05/03 13:46:35	f582e064be47	1	样本数据	确认	http://camaradeideias.com.br/...	<input type="button" value="标记"/> <input type="button" value="备注"/> <input type="button" value="推送"/>
566202	2018/05/01 07:45:16	f582e064be47	1	样本数据	确认	http://camaradeideias.com.br/...	<input type="button" value="标记"/> <input type="button" value="备注"/> <input type="button" value="推送"/>
566039	2018/04/30 21:01:20	f582e064be47	1	样本数据	确认	http://camaradeideias.com.br/...	<input type="button" value="标记"/> <input type="button" value="备注"/> <input type="button" value="推送"/>

```
C:\Windows\system32\cmd.exe
CDrive.exe           4576 Console           1      66,780 K
ONENOTEM.EXE         4888 Console           1       3,816 K
chrone.exe            21776 Console          1     29,800 K
end.exe               38216 Console          1       4,804 K
findstr.exe           38252 Console          1       3,808 K

d:\Users\deng_hy>netstat -ano | findstr "80"
TCP        [REDACTED]:50717          [REDACTED] ESTABLISHED 24924
TCP        [REDACTED]:50718          [REDACTED] ESTABLISHED 24924
TCP        [REDACTED]:50719          [REDACTED] ESTABLISHED 24924
TCP        [REDACTED]:50805          [REDACTED] ESTABLISHED 2276
TCP        [REDACTED]:50973          109.230.199.169:80 SYN_SENT    4396
UDP        127.0.0.1:54054          *:*          480
UDP        [fe80::9409:9745:6c:1b6z12]:546 *:*          916
UDP        [fe80::9409:9745:6c:1b6z12]:1900 *:*          3540
UDP        [fe80::9409:9745:6c:1b6z12]:54040 *:*          3540

d:\Users\deng_hy>tasklist | findstr "4396"
21648976.exe           4396 Console           1       7,328 K

d:\Users\deng_hy>
```



ML&恶意域名检测——后续改进

误报


<http://www.jxskqyy.com/>

The screenshot shows a web browser displaying the homepage of the Affiliated Stomatological Hospital of Nanchang University. The browser's address bar shows the URL www.jxskqyy.com. The website header includes a navigation menu with links for 'OA 办公系统', '收藏本站', and '设为首页'. The main content area features the hospital's logo, name in Chinese and English, and contact information. A search bar is present with the placeholder text '请输入关键字!' and a '搜索' button. Below the search bar, there is a list of popular search terms: '热门搜索: 假牙 牙齿矫正 植牙 补牙 洗牙'. At the bottom of the page, there is a blue navigation bar with links for '医院概况', '在线挂号', '专科介绍', '就医指南', '口腔保健', '新闻中心', '科研管理', '教学园地', '下载中心', '专题', and '口腔医学院'.



2018 携程安全沙龙

ML&恶意域名检测——API校验注意点













One engine detected this URL

URL <http://baidu.com/>
Host baidu.com

Downloaded file 1d2d898b8ee2611eeed02868e27b1e3db2d266b4016b54160f128187961431fb
Last analysis 2018-09-01 15:25:08 UTC
Community score -129

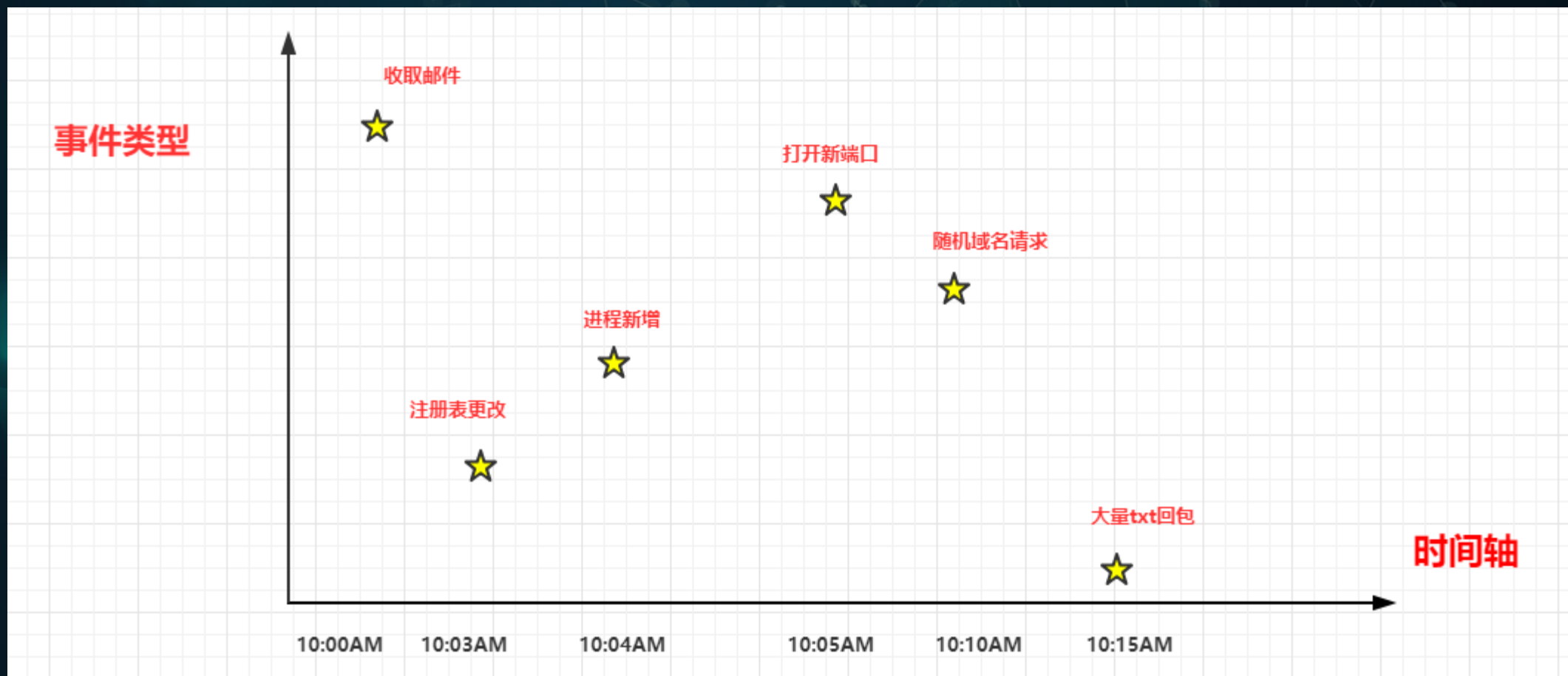
1 / 67

Detection Details Community 3

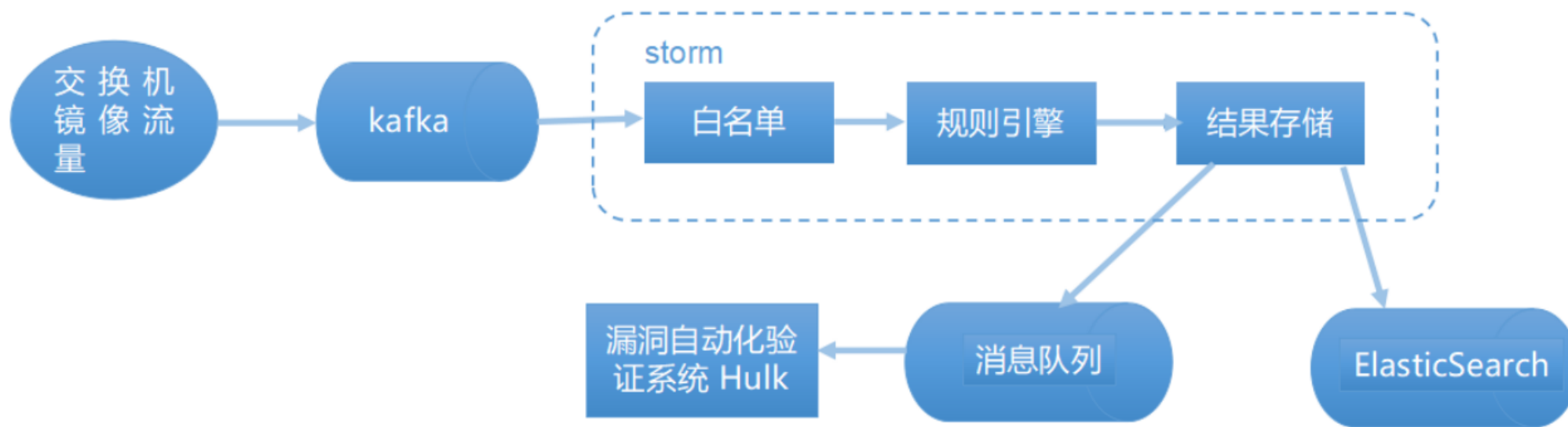
CLEAN MX	 Phishing	Quttera	 Suspicious
ADMINUSLabs	 Clean	AegisLab WebGuard	 Clean
AlienVault	 Clean	Antiy-AVL	 Clean
Avira	 Clean	Baidu-International	 Clean
BitDefender	 Clean	Blueliv	 Clean



ML&恶意域名检测——产品的下一步融合



ML&web攻击检测——攻击复验系统介绍



ML&web攻击检测——攻击复验系统介绍

Storm 实时大数据

结果列表 规则列表 操作日志 黑/白名单 攻击事件 规则字典管理 使用帮助

结束时间 11/17/2017 00:00 去重 查询 导出去重结果

ID	发现时间 ^	攻击类型	源IP	Count	Method	URL
AV_Cutxl1H046vBwkZel	2017-11-16 10:49:05	XXE-1010	27.151.112.217	1	POST	████████ctrip.com/████████/http
AV_Cuge21H046vBwkBfb	2017-11-16 10:48:11	XXE-1010	27.151.112.217	1	POST	████████ctrip.com/████████http
AV_Cufoe1H046vBwkAEU	2017-11-16 10:48:08	XXE-1010	27.151.112.217	1	POST	████████ctrip.com,████████http
AV_CugHs1H046vBwkAZy	2017-11-16 10:48:08	InfoLeak-1006	27.151.112.217	1	GET	████████ctrip.com/.bash_history
AV_CufjNv2KjZyxKTZNm	2017-11-16 10:48:06	RFI-1004	27.151.112.217	1	GET	████████ctrip.com/████████sh?HOSTSVC=../../../../etc/passwd
AV_CufjNv2KjZyxKTZNg	2017-11-16 10:48:05	InfoLeak-1006	27.151.112.217	1	GET	████████ctrip.com/████████/site.ini

请求头信息

请求 :

GET : █████████ctrip.com/████████/site.ini

Host : █████████ctrip.com

PostData :

Referer :

User-Agent : Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0)

响应 :

Status-Code : 430

来源 : 福建福州



ML&web攻击检测

为什么想到用机器学习来做web攻击检测？

例如一条检测sql注入的正则语句如下：

```
String inj_str =  
"|and|exec|insert|select|delete|update|co  
unt|*|%|chr|mid|master|truncate|char|decl  
are|;|or|-|+|,";
```

新买的selected衬衫脏了!



ML&web攻击检测

为什么想到用机器学习来做web攻击检测？

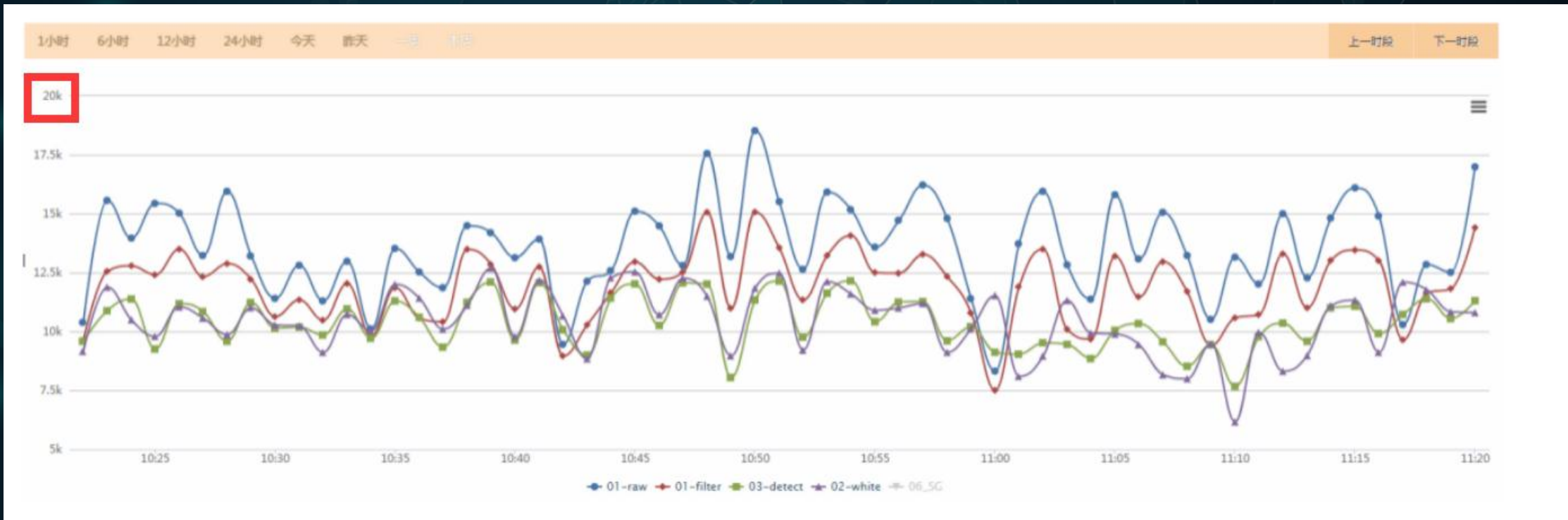
pattern:[^\w\s\?此处保密!\@\(\|\~]{1,}

规则难以维护，自己写的正则自己都读不懂



ML&web攻击检测——未改进前

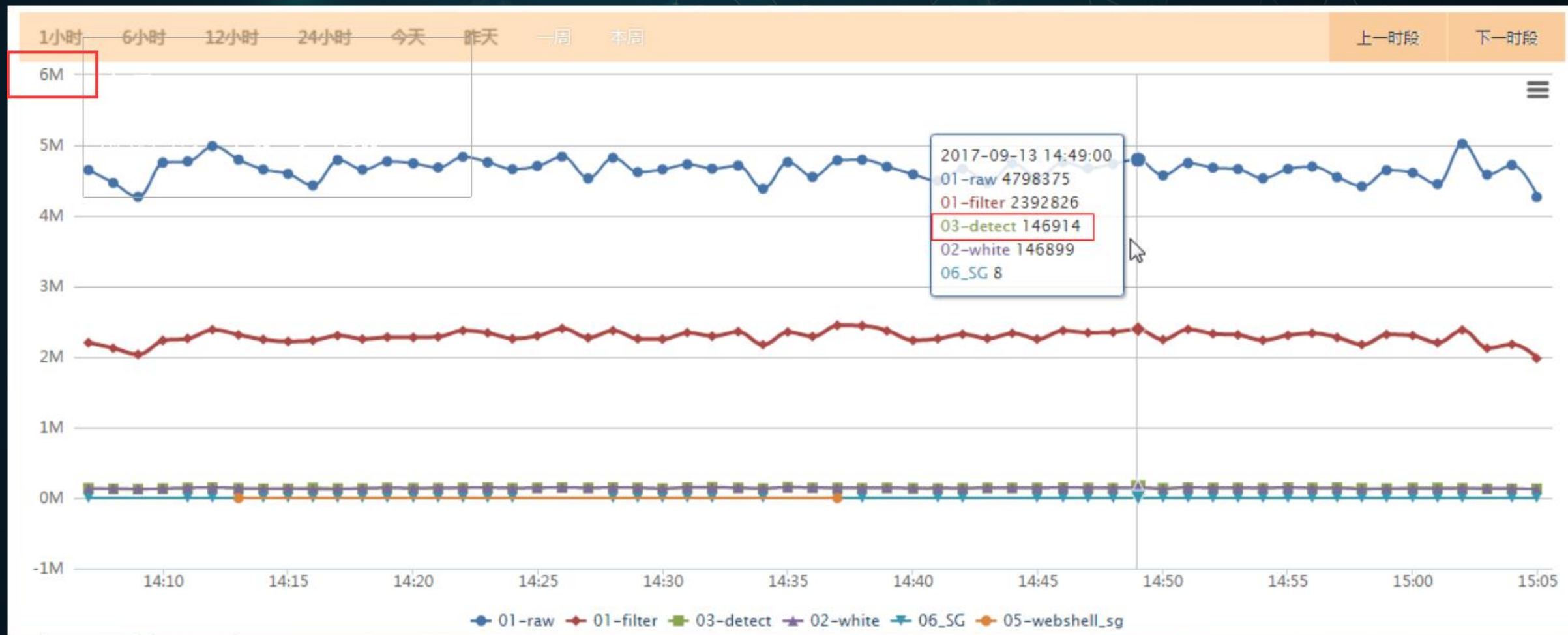
为什么想到用机器学习来做web攻击检测？



队列积压严重，根本消费不完，昨天的攻击今天还没检测



ML&web攻击检测——改进后性能效果

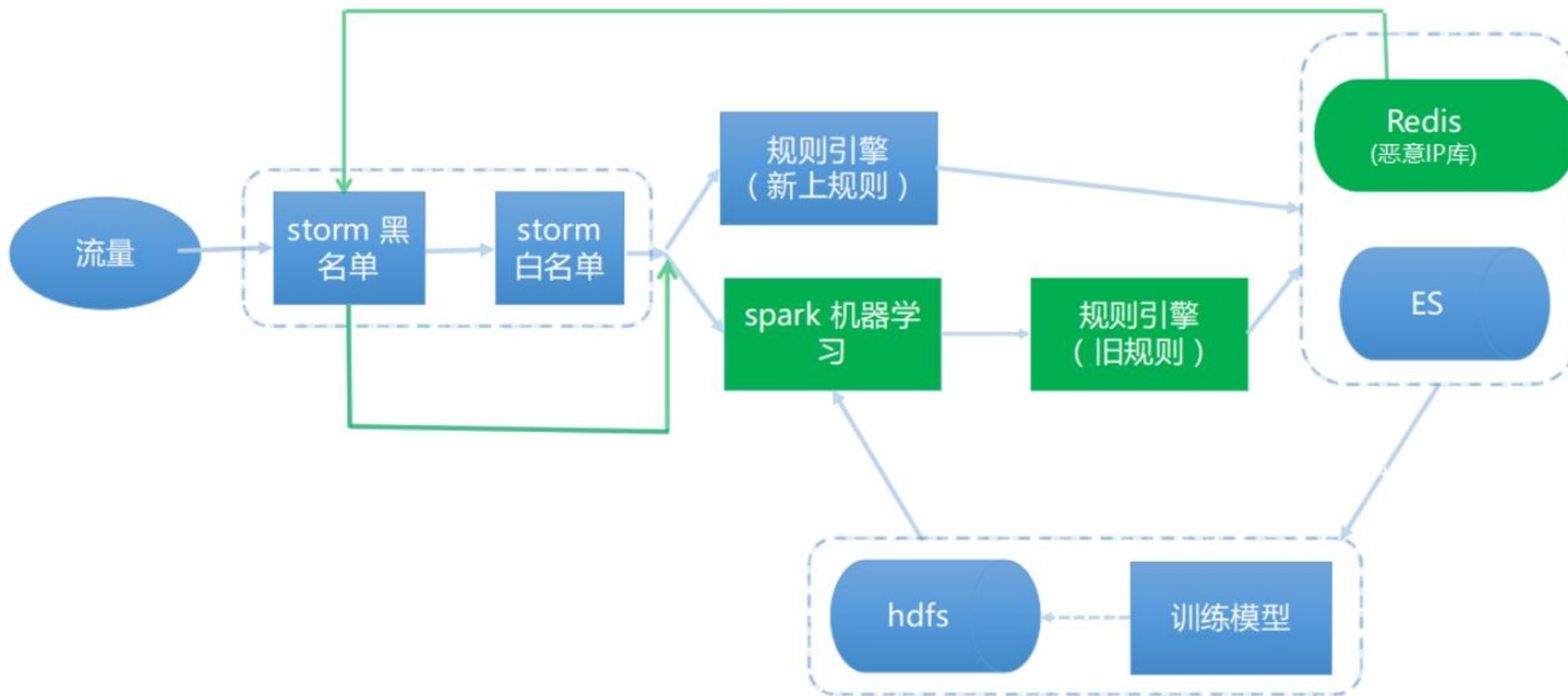


ML&web攻击检测——和正则对比效果

Time	is_ml	rule_result	postdata
▶ July 3rd 2018, 15:34:04.669	true	white	out_trade_no=1122%00'&subject=0
▶ July 3rd 2018, 15:34:01.091	true	white	Y2hlbmdzaGlzMjd=echo md5(zhimakaimen233333333);
▶ July 3rd 2018, 15:33:02.170	true	white	second=1@%7Cecho%20%27%3C%3Fphp%20phpinfo%28%29%3B%3F%3E%27%20%3E%20dsanhsds.php
▶ July 3rd 2018, 15:32:01.845	true	white	name=1@ echo '<?php phpinfo();?>' > ssfds.php
▶ July 3rd 2018, 15:28:07.650	true	white	{"username":"" and 1=char(106) --"}
▶ July 3rd 2018, 15:21:07.303	true	white	key=%' and 1=char(106)-- &method=GetDeptJSON&org=2&all=0&hr=0&d11=IBP_Core.d11&c1ass=IBP_Core.OrgStructure
▶ July 3rd 2018, 15:21:04.452	true	white	uid`%3d-42873%0bor%0b42873%3d42873%23=1
▶ July 3rd 2018, 15:21:01.360	true	white	login=%df%27test&email=test@test.com
▶ July 3rd 2018, 15:16:03.862	true	white	username=eye' and 1=char(106) --&password=sdasd&expires=&butOk=



ML&web攻击检测——第6版框架



ML&web攻击检测——模块解释

args参数值不带英文标点和控制字符的请求，全都算白名单，不过任何检测引擎

www.ctrip.com?目的地=北京

白名单

过完白名单的请求，进入ML引擎，ML预测为黑则继续进入正则，
否则打上ML白色结果

ML和正则引擎
的关系

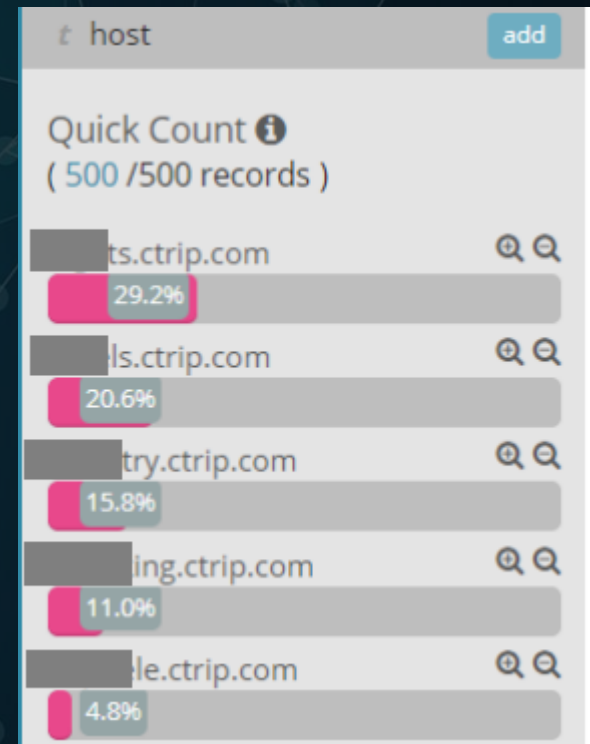
命中正则规则超过20次的ip，进入黑名单，缓存1day，此ip之后不过ML引擎，
直接进正则

恶意IP库



ML&web攻击检测——聊下训练数据的搜集

白日志来源



难点：

- 1.大部分都是酒店机票的日志
- 2.大部分流量都不带任何标点和特殊字符
- 3.要达到黑白样本1:1

写脚本ES捞，手工洗，去重，日志尽量多样化
针对样本污染问题，使用正则关键字从白样本里洗掉黑色数据



01 GET ,POST分开 取args进行建模

<http://test.ctrip.com/TrainBooking/Search.aspx?from=shanghai&to=beijing&day=2018-09-05>

按args取训练素材，还方便使用网上直接搜集来的POC，因为不用考虑定制化

过拟合



ML&web攻击检测——Details过拟合

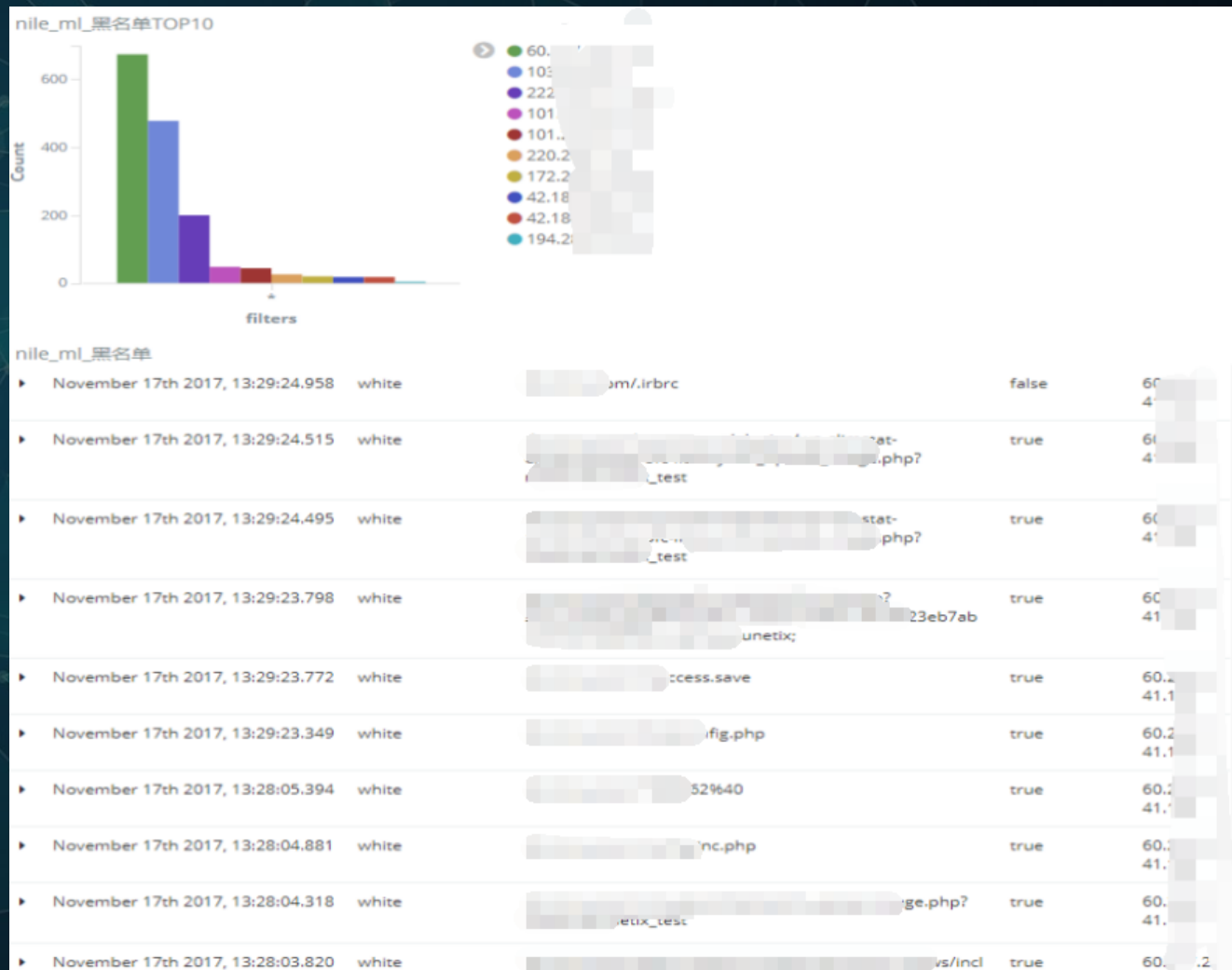


02 图片上传，加密数据的怎么办？

**正则检测不了的，机器学习也不能强求
多层防御，用其他方式来解决**



03 漏报了怎么办



04 误报了怎么办

误报的影响不大，直接交给正则，不同于WAF的使用场景

使用HMM做异常检测



HMM异常检测——问题抛出

世界杯开始前，C罗状态好的概率为90%

.....

C罗状态好，葡萄牙赢球概率为90%

.....

C罗这一场状态好，下一场状态好的概率为85%

....

问葡萄牙世界杯开始3连跪的概率是多少

■ 选择	门将	1	❄️ 卡西利亚斯
■ 换人	中后卫	6	➡️ 伊万.埃尔格拉
■ 玩家	中后卫	19	➡️ 萨穆埃尔
■ 镜头	右边卫	2	➡️ 萨尔加多
■ 画面	左边卫	3	➡️ 罗伯特.卡洛斯
■ 声音	后腰	16	➡️ 格拉维森
■ 按钮	中前卫	23	➡️ 贝克汉姆
■ 指令	左前卫	10	➕ 费戈



HMM异常检测——泛型化

某参数正常的张这样，问题：来一个请求，判断是否是异常的

ark_bus_vivo|12308

ark_bus_xiaomi|12308

ark_android_jpskb|TY

ark_bus_hicloud|ky12308

首先做泛型：

字母-> ord(A) 数字-> ord('N') 中文-> ord('C') 其他-> 取其ASCII码

a r k _ | 1

原始序列

65 65 65 95 124 49

泛化后的观察序列

S1 S1 S1 S2 S2 S3

隐藏序列



HMM异常检测——计算观察序列概率

fromCity=

正常请求	恶意请求
● shanghai 3分	● <img src=1 -100分
● xi'an 1分	● select 1,2,3 -30分
● Los Angels 4分	● etc/passwd -50分
●	●
● St. Paul 2分	● 127.0.0.1&&w hoami -500分

学习正常请求，算出最“异常”的一个正常请求的评分，作为“理论”临界点



HMM异常检测——样本里的异常点和分布

deviceId
00277AD2-AFD1-4F79-AED9-E41...
00000000-0000-0000-0000-000000000000
001A7481-9EC4-42AB-97FA-7...
005325AA8D278EB3C5DC...

deviceId
00000000-0000-0000-0000-0000-000000000000得分：
-114.04052019890963

<script>alert(123)得分：
43.74534140982247

样本收集最好做到分散：从不同的源ip收集，避免单个ip贡献过多样本，恶意ip库里面的数据坚决删除

监测模型是否待更新？或者干脆定期更新



Questions





THANKS



2018 携程安全沙龙