

实时攻击检测智能化之路

携程信息安全部 – 陈莹

CONTENTS

● 系统介绍

● 智能进化

● 未来

系统介绍 - Nile



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	knair.flights.ctrip.com/lcds/messagebroker/http
AV_Cuge21H046vBwkBfb	2017-11-16 10:48:11	XXE-1010	27.151.112.217	1	POST	knair.flights.ctrip.com/messagebroker/http
AV_Cufoe1H046vBwkAEU	2017-11-16 10:48:08	XXE-1010	27.151.112.217	1	POST	knair.flights.ctrip.com/lcds/messagebroker/http
AV_CugHs1H046vBwkA2y	2017-11-16 10:48:08	InfoLeak-1006	27.151.112.217	1	GET	knair.flights.ctrip.com/. bash_history
AV_CufJNv2KjZyxKTZNm	2017-11-16 10:48:06	RFI-1004	27.151.112.217	1	GET	knair.flights.ctrip.com/scripts/bb-hostsvc.sh?HOSTSVC=../../../../etc/passwd
AV_CufJNv2KjZyxKTZNg	2017-11-16 10:48:05	InfoLeak-1006	27.151.112.217	1	GET	knair.flights.ctrip.com/scripts/settings/site. ini

请求头信息

请求：

GET : knair.flights.ctrip.com/scripts/settings/site.ini

Host : knair.flights.ctrip.com

PostData：

Referer：

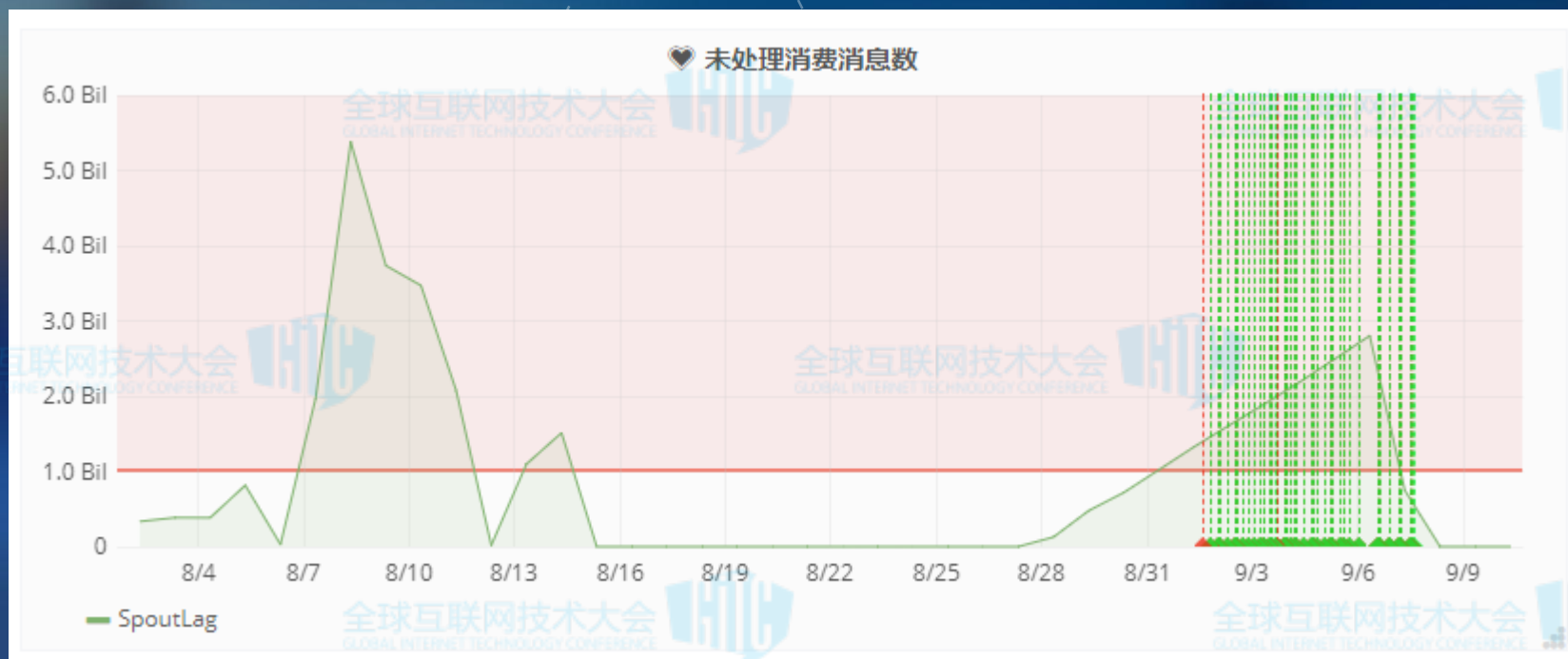
User-Agent : Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 5.1; Trident/4.0)

响应：

Status-Code : 430

来源：福建福州

存在的问题



机器学习可以解决的问题



方案选择

机器学习

Python

spark
mllib

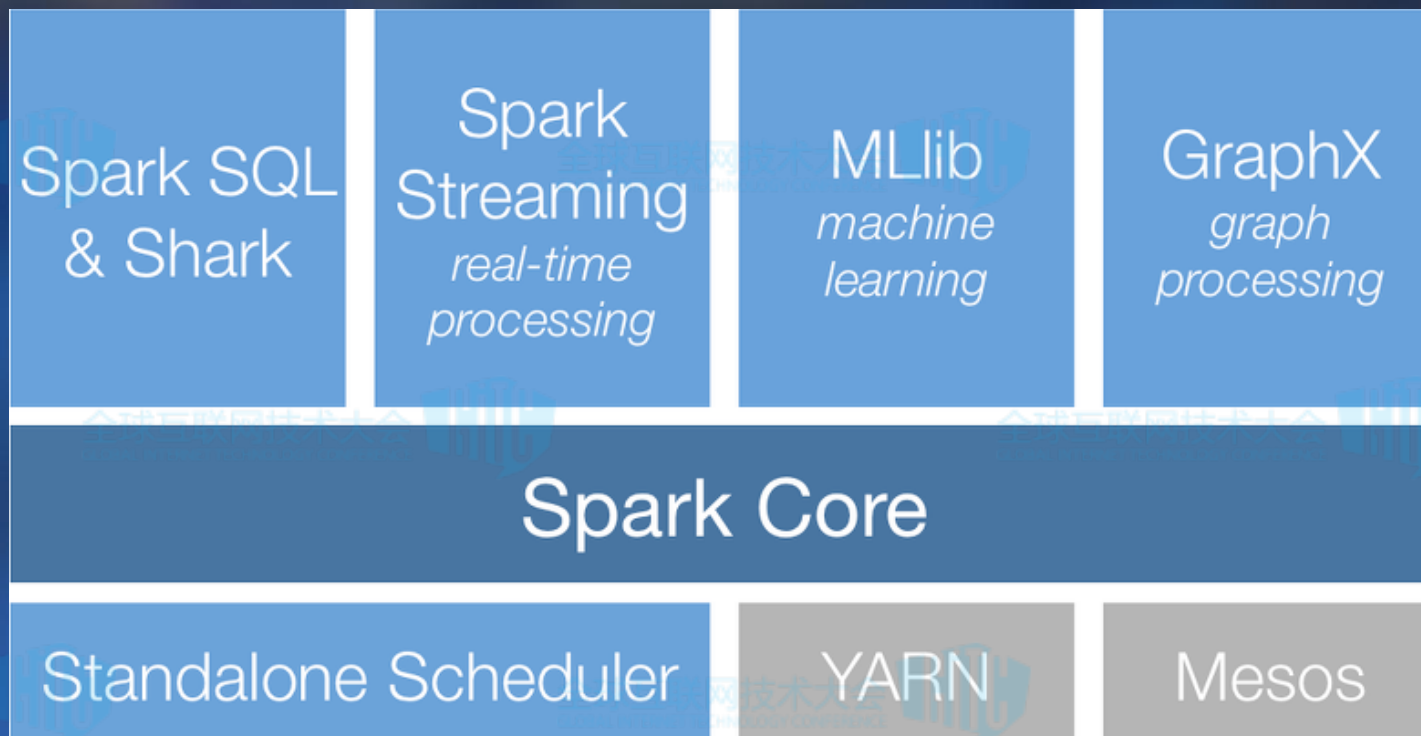
java

scala

sckit-learn

Spark 介绍

Apache Spark™ is a fast and general engine for large-scale data processing.



机器学习流程

1. 收集样本
2. 数据清洗，打标签
3. 特征提取
4. 模型训练
5. 验证模型，调优
6. 预测分类

1.0 之架构



1.0

- 数据：url和postdata
- 特征选择：url decode+正则
- 算法：svm
- 算法库：spark mllib

1.0 之样本

样本收集

恶意样本

- nile 命中规则的结果
- waf 日志
- 网上收集poc

非恶意样本

- 交换机镜像的流量

样本清洗

脚本关键字+ 人眼观察

1.0 之特征

统计每个request中如下敏感符号,关键字的个数

,.!*\\/(& < > 等等

eval ongl script select等等

然后转换成一个 $1 \times n$ 的矩阵,所有的训练样本就是 $m \times n$ 的输入

and 1=(select count(*) from master.dbo.sysobjects where xtype = 'x' and name = 'xp_cmdshell')

[2,1,3,.....1,2,0] , 代表2个(,2个),3个'等

存在的问题:

总有遗漏的关键词

1.0 之算法

测试算法	误报率	漏报率
决策树	9.9%	8.4%
svm	8.8%	8.9%
朴素贝叶斯	11%	9.6%

1.0

全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



进步

从无到有，流程跑通

不足

特征太依赖于正则了，
不够智能

全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



2.0

- 分词：WordParser
- 特征提取：TF-IDF (Hashing TF and IDF)

特征之WordParser

全球互联网技术大会



全球互联网技术大会



将每一个标点和控制符都“转换”为词，例如

and 1=(select count(*) from master.dbo.sysobjects where xtype =
'x' and name = 'xp_cmdshell')

```
om master.dbo.sysobjects where xtype = 'x' and name = 'xp_cmdshell') "
```

最大化

```
and  
1  
=  
(  
select  
count  
(  
*  
)  
from  
master.dbo.sysobjects  
where  
xtype  
=  
'  
x'  
,  
and  
name  
='xp_cmdshell'  
)
```

特征提取之TF-IDF

$$\text{词频(TF)} = \frac{\text{某个词在文章中的出现次数}}{\text{文章的总词数}}$$

$$\text{逆文档频率(IDF)} = \log\left(\frac{\text{语料库的文档总数}}{\text{包含该词的文档数} + 1}\right)$$

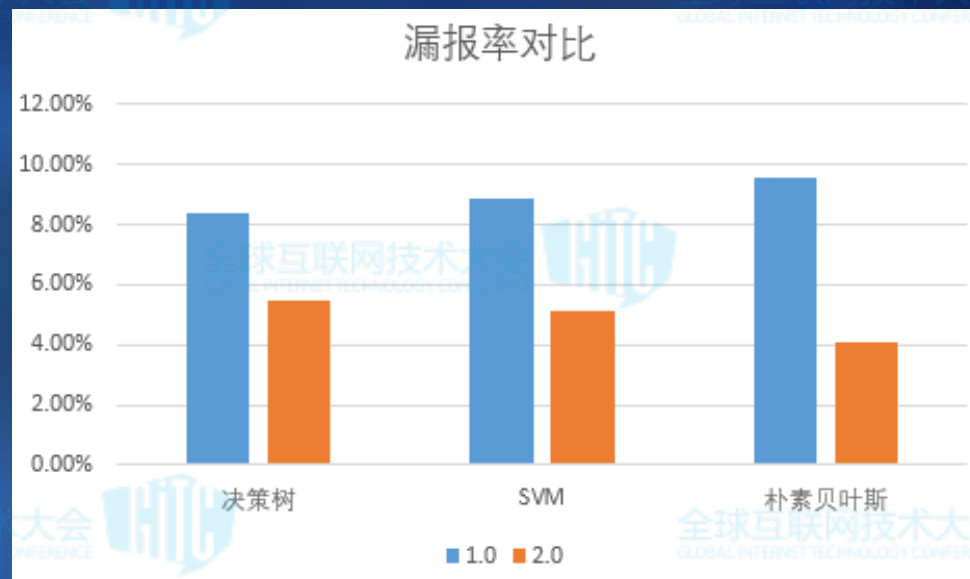
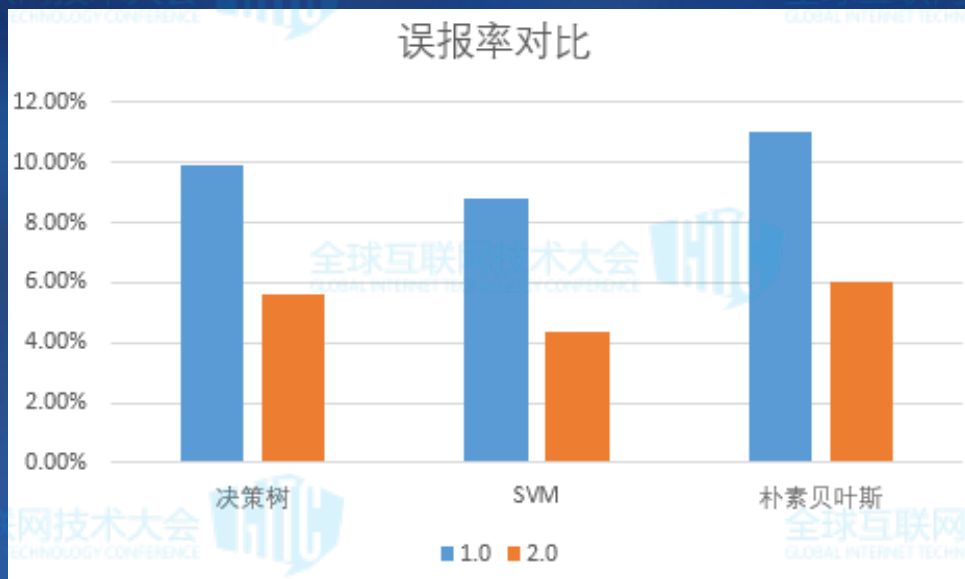
$$\text{TF-IDF} = \text{词频(TF)} \times \text{逆文档频率(IDF)}$$

例如我们有很多条get请求语句，第一条语句共计10个单词，其中单引号有3个，1000条语句中有10条语句包含单引号

	包含该词的语句个数	TF	IDF	TF-IDF
单引号	10	0.3	1.958	0.5874
from	100	0.3	0.995	0.3318

2.0 之 算法

测试算法	误报率	漏报率
决策树	5.6%	5.5%
svm	4.4%	5.1%
朴素贝叶斯	6.0%	4.1%



效果

rule_result	url	postdata
white	hotels.ctrip.com:80/hotel/4688225.html	(%23_memberAccess%3d@ognl.OgnlContext@DEFAULT_MEMBER_ACCESS)%3f(%23req%3d%40org.apache.struts2.ServletActionContext%40getRequest(),%23res%3d%40org.apache.struts2.ServletActionContext%40getResponse(),%23res.setCharacterEncoding(%23parameters.encoding[0]),%23w%3d%23res.getWriter(),%23w.print(%23parameters.web[0]),%23w.print(%23parameters.path[0]),%23w.close()):xx.toString.json?&pp=%2f&encoding=UTF-8&web=security_&path=check
white	you.ctrip.com:80/sightlist/chungcheongnam1445.html	-----7e116d19044c Content-Disposition: form-data; name="test"; filename="%{(#test='multipart/form-data').(#dm=@ognl.OgnlContext@DEFAULT_MEMBER_ACCESS).(#_memberAccess?(#_memberAccess=#dm):((#container=#context['com.opensymphony.xwork2.ActionContext.container']).(#ognlUtil=#container.getInstance(@com.opensymphony.xwork2.ognl.OgnlUtil@class)).(#ognlUtil.getExcludedPackageNames().clear()).(#ognlUtil.getExcludedClasses().clear()).(#context.setMemberAccess(#dm))))).(#req=@org.apache.struts2.ServletActionContext@getRequest()).(#res=@org.apache.struts2.ServletActionContext@getResponse()).(#res.setContentType('text/html;charset=UTF-8')).(#res.getWriter().print('security_')).(#res.getWriter().print('check')).(#res.getWriter().flush()).(#res.getWriter().close())}b"
white	you.ctrip.com:80/sightlist/chungcheongnam1445.html	method:%23_memberAccess%3d@ognl.OgnlContext@DEFAULT_MEMBER_ACCESS,%23req%3d%40org.apache.struts2.ServletActionContext%40getRequest(),%23res%3d%40org.apache.struts2.ServletActionContext%40getResponse(),%23res.setCharacterEncoding(%23parameters.encoding[0]),%23w%3d%23res.getWriter(),%23w.print(%23parameters.web[0]),%23w.print(%23parameters.path[0]),%23w.close(),1?%23xx:%23request.toString&pp=%2f&encoding=UTF-8&web=security_&path=check
white	you.ctrip.com:80/sightlist/chungcheongnam1445.html	debug=command&expression=%23req%3d%23context.get(%27co%27%2b%27m.open%27%2b%27symphony.xwo%27%2b%27rk2.disp%27%2b%27atcher.HttpSer%27%2b%27vletReq%27%2b%27uest%27),%23resp%3d%23context.get(%27co%27%2b%27m.open%27%2b%27symphony.xwo%27%2b%27rk2.disp%27%2b%27atcher.HttpSer%27%2b%27vletRes%27%2b%27ponse%27),%23resp.setCharacterEncoding(%27UTF-8%27),%23resp.getWriter().print(%22security_%22),%23resp.getWriter().print(%22check%22),%23resp.getWriter().flush(),%23resp.getWriter().close()

效果

rule_result	url
white	zhair.flights.ctrip.com/images/upload/fckimg//editor/filemanager/browser/default/connectors/aspx/connector.aspx? Command=GetFolders&Type=File&CurrentFolder=%2F
white	zhair.flights.ctrip.com/editor1//editor/filemanager/browser/default/connectors/aspx/connector.aspx? Command=GetFolders&Type=File&CurrentFolder=%2F
white	zhair.flights.ctrip.com/admin/fck//editor/filemanager/connectors/aspx/connector.aspx? Command=GetFolders&Type=File&CurrentFolder=%2F
white	zhair.flights.ctrip.com/includes/fckeditor/editor/filemanager/connectors/aspx/upload.aspx? Command=CreateFolder&Type=Media&CurrentFolder=ali.asp&NewFolderName=hack.asp
white	zhair.flights.ctrip.com/manage/fckeditor//editor/filemanager/connectors/aspx/connector.aspx? Command=GetFolders&Type=File&CurrentFolder=%2F
white	zhair.flights.ctrip.com/includes/fckeditor/editor/filemanager/browser/default/connectors/aspx/connector.aspx? Command=CreateFolder&Type=Media&CurrentFolder=ali.asp&NewFolderName=hack.asp
white	zhair.flights.ctrip.com/fckeditor/editor/filemanager/connectors/aspx/connector.aspx? Command=CreateFolder&Type=Media&CurrentFolder=ali.asp&NewFolderName=hack.asp
white	zhair.flights.ctrip.com/images/upload/fckeditor//editor/filemanager/browser/default/connectors/aspx/connector.aspx? Command=GetFolders&Type=File&CurrentFolder=%2F
white	zhair.flights.ctrip.com/include/fckeditor//editor/filemanager/connectors/aspx/connector.aspx? Command=GetFolders&Type=File&CurrentFolder=%2F

2.0

进步
可以靠数据变得更强
真正开始智能化

不足
未解决性能问题

3.0

全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



•架构调整

全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



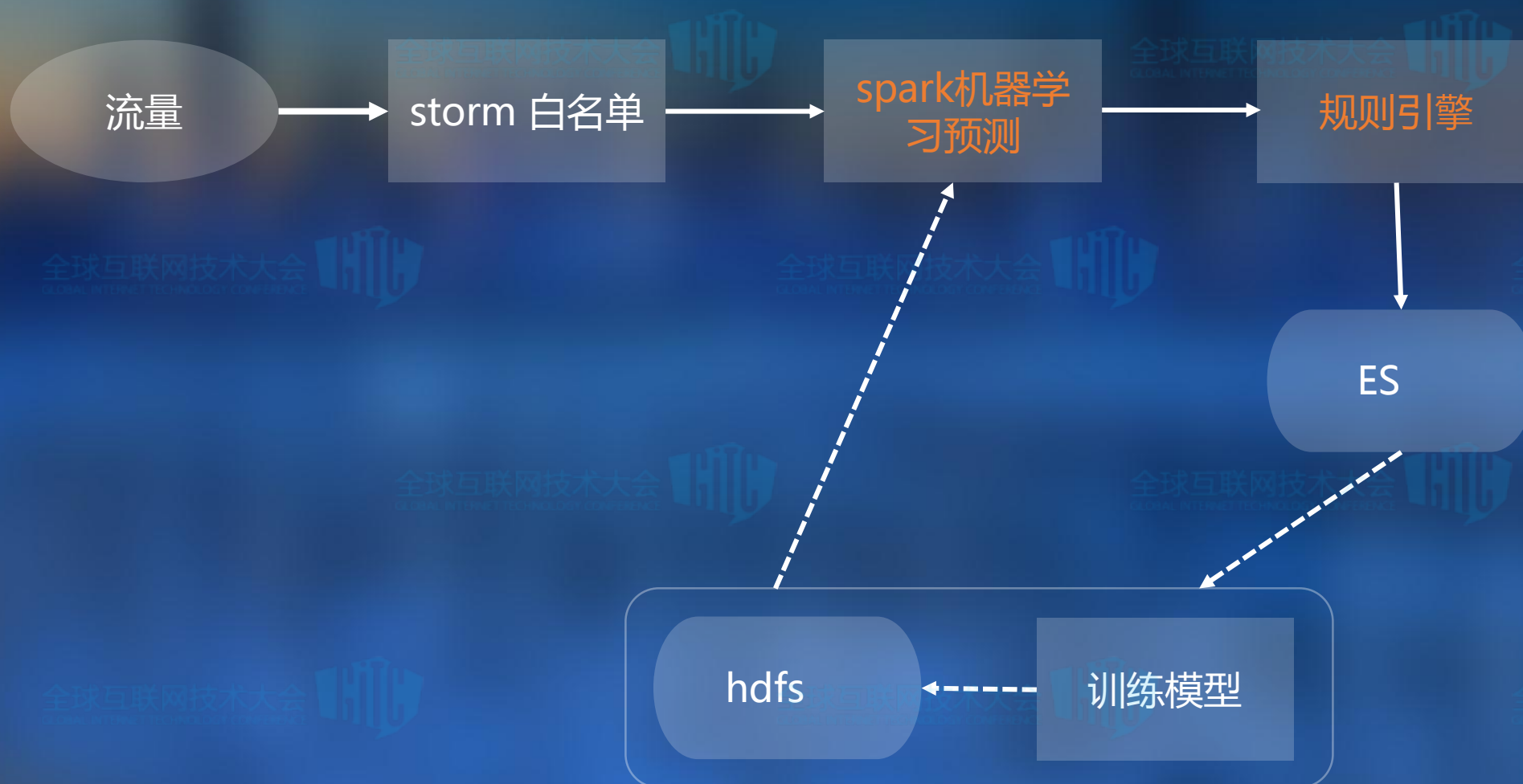
全球互联网技术大会



全球互联网技术大会



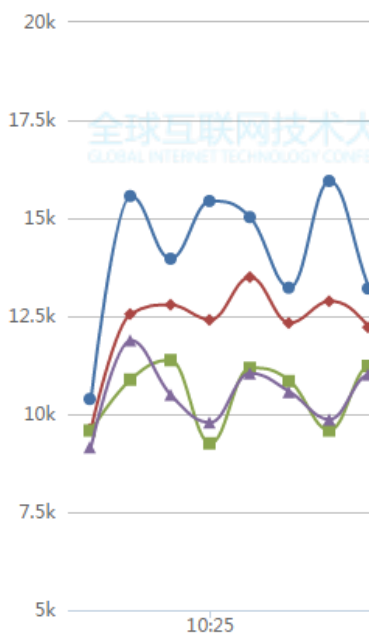
3.0 -- 架构



性能效果

1小时流量: 1.72k

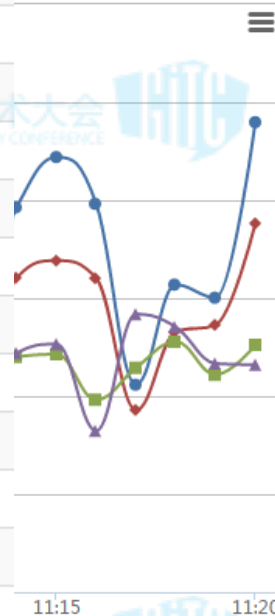
1小时 6小时 12小时 24小时



Batch Time	Input Size	Scheduling Delay (?)	Processing Time (s)
2017/09/13 15:17:00	150060 records	1 ms	12 s
2017/09/13 15:16:00	141986 records	0 ms	12 s
2017/09/13 15:15:00	139310 records	1 ms	11 s
2017/09/13 15:14:00	143154 records	0 ms	14 s
2017/09/13 15:13:00	141412 records	0 ms	11 s
2017/09/13 15:12:00	144183 records	0 ms	9 s
2017/09/13 15:11:00	138273 records	1 ms	9 s
2017/09/13 15:10:00	136210 records	1 ms	11 s
2017/09/13 15:09:00	136938 records	1 ms	13 s
2017/09/13 15:08:00	137362 records	0 ms	11 s
2017/09/13 15:07:00	138267 records	1 ms	11 s
2017/09/13 15:06:00	139190 records	0 ms	12 s
2017/09/13 15:05:00	137216 records	1 ms	11 s
2017/09/13 15:04:00	135794 records	0 ms	12 s
2017/09/13 15:03:00	138136 records	1 ms	8 s
2017/09/13 15:02:00	139489 records	0 ms	10 s

上一时段 下一时段

上一时段 下一时段



3.0

进步

性能大幅提高

不足

如果新上规则的话，
很大概率检测不出来

4.0

全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



•再次调整架构

全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



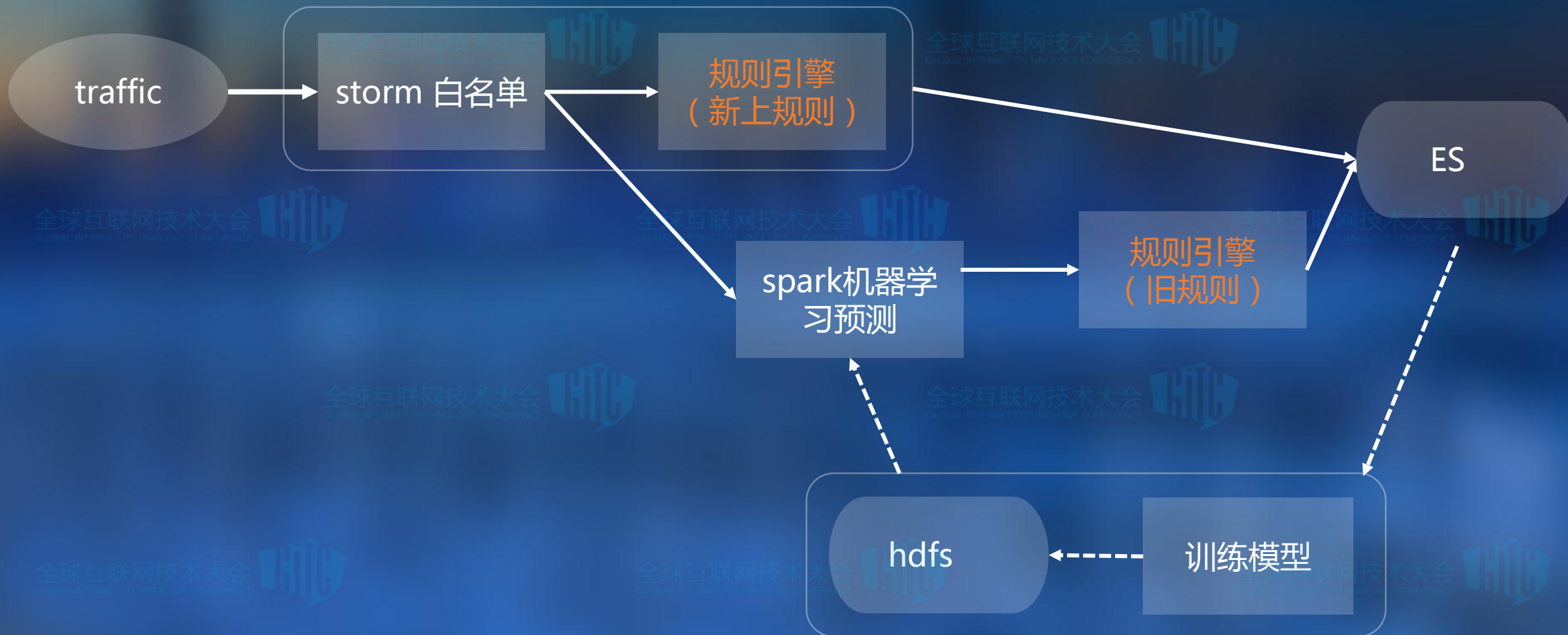
全球互联网技术大会



全球互联网技术大会



4.0 -- 架构



4.0

进步

可抓新规则定义的攻
击

不足

无法检测header头

5.0

全球互联网技术大会



全球互联网技术大会



•增加动态黑名单功能

全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会

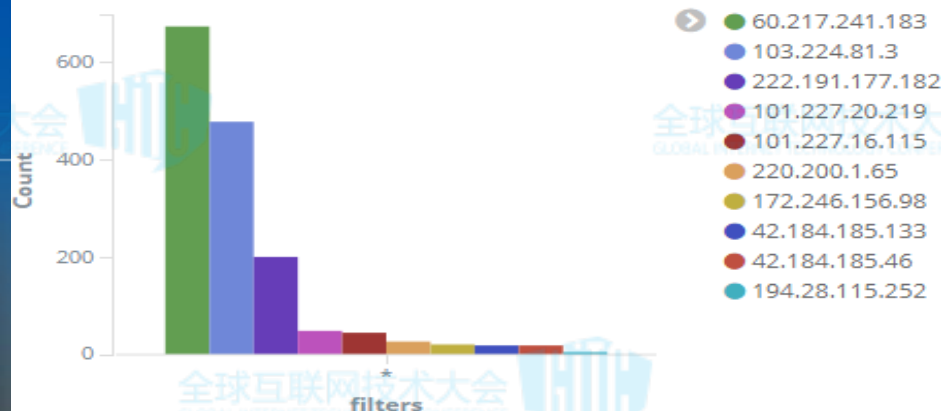


5.0 - 效果

t hit_rule	QQ	* SQLi-0205		
t host	QQ	* ru.ctrip.com		
is_black	QQ	* true		
t method	QQ	t hit_rule	QQ	* SQLi-0205
t port	QQ	t host	QQ	* es.ctrip.com
t postdata	QQ	is_black	QQ	* true
t referer	QQ	t method	QQ	* GET
t rule_result	QQ	t port	QQ	* 80
t sip	QQ	t postdata	QQ	*
t status	QQ	t referer	QQ	* 1 waitfor delay '0:0:9' --
t uagent	QQ	t rule_result	QQ	* black
t uri	QQ	t sip	QQ	* 60.217.241.183
t url	QQ	t status	QQ	* 200
		t uagent	QQ	* Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.21 (KHTML, like Gecko) Chrome/41.0.2228.0 Safari/537.21
		t uri	QQ	* /hotels/lijiang-hotels-list-37/
		t url	QQ	* es.ctrip.com/hotels/lijiang-hotels-list-37/

November 14th 2017, 19:37:06.503	white	es.ctrip.com/hotels/guangzhou-hotels-list-32/	true	200	60.217. 241.183
November 14th 2017, 19:37:06.502	white	jp.ctrip.com/hotels/bali-hotels-list-723/	true	200	101.227 .16.115
November 14th 2017, 19:37:06.500	white	jp.ctrip.com/hotels/xi-an-hotels-list-10/	true	200	101.227 .16.115
November 14th 2017, 19:37:06.500	white	de.ctrip.com/hotels/macau-hotels-list-59/	true	200	101.227 .16.115
November 14th 2017, 19:37:06.492	white	es.ctrip.com/hotels/lijiang-hotels-list-37/	true	200	60.217. 241.183
November 14th 2017, 19:37:06.485	white	de.ctrip.com/hotels/macau-hotels-list-59/	true	200	101.227 .16.115
November 14th 2017, 19:37:06.480	white	jp.ctrip.com/hotels/wuhan-hotels-list-477/	true	200	101.227 .16.115
November 14th 2017, 19:37:06.479	white	de.ctrip.com/hotels/macau-hotels-list-59/	true	200	101.227 .16.115
November 14th 2017, 19:37:06.477	white	jp.ctrip.com/hotels/lijiang-hotels-list-37/	true	200	101.227 .16.115
November 14th 2017, 19:37:06.475	black	ru.ctrip.com/hotels/lijiang-hotels-list-37/	true	200	101.227 .20.219
November 14th 2017, 19:37:06.469	white	fr.ctrip.com/hotels/hong-kong-hotels-list-58/	true	200	60.217. 241.183
November 14th 2017, 19:37:06.468	white	ru.ctrip.com/hotels/ningbo-hotels-list-375/	true	200	101.227 .20.219
November 14th 2017, 19:37:06.468	white	es.ctrip.com/hotels/guangzhou-hotels-list-32/	true	200	60.217. 241.183

nile_ml_黑名单TOP10



nile_ml_黑名单

▶	November 17th 2017, 13:29:24.958	white	dst.ctrip.com/.irbrc	false	60.217.241.183
▶	November 17th 2017, 13:29:24.515	white	dst.ctrip.com/wp-content/plugins/wp-slimstat-ex/lib/ofc/php-ofc-library/ofc_upload_image.php?name=acunetix_test	true	60.217.241.183
▶	November 17th 2017, 13:29:24.495	white	dst.ctrip.com/wp-content/plugins/wp-slimstat-ex/lib/ofc/php-ofc-library/ofc_upload_image.php?name=acunetix_test	true	60.217.241.183
▶	November 17th 2017, 13:29:23.798	white	dst.ctrip.com/lib/phpThumb/phpThumb.php?src=./index.php&fltr[]=blur 5;echo+082119f75623eb7abd7bf357698ff66c>cache/acunetix;	true	60.217.241.183
▶	November 17th 2017, 13:29:23.772	white	dst.ctrip.com/.htaccess.save	true	60.217.241.183
▶	November 17th 2017, 13:29:23.349	white	dst.ctrip.com//localconfig.php	true	60.217.241.183
▶	November 17th 2017, 13:28:05.394	white	dst.ctrip.com//'"908062%40	true	60.217.241.183
▶	November 17th 2017, 13:28:04.881	white	dst.ctrip.com//config.inc.php	true	60.217.241.183
▶	November 17th 2017, 13:28:04.318	white	dst.ctrip.com/includes/ofc/php/ofc_upload_image.php?name=acunetix_test	true	60.217.241.183
▶	November 17th 2017, 13:28:03.820	white	dst.ctrip.com/administrator/components/com_jnews/incl	true	60.217.241.183

5.0

全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



进步

更低的漏报
可发现业务逻辑攻击

不足

全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



全球互联网技术大会



未来



要解决的问题

- post数据中xml和json格式的数据存在大量误报



更智能

- 二分类到多分类
- 使用更多的检测纬度
- 关联各个纬度



增加反馈

每周报告效果

全球互联网技术大会
GLOBAL INTERNET TECHNOLOGY CONFERENCE



THANKS

更多细节请关注公众号“携程技术中心”，会整理文字版投稿上去