
Alibaba Group
阿里巴巴集团
×
蚂蚁金服
ANT FINANCIAL
“业务为王”时代下的WAF新趋势

企业可能遇到的危险

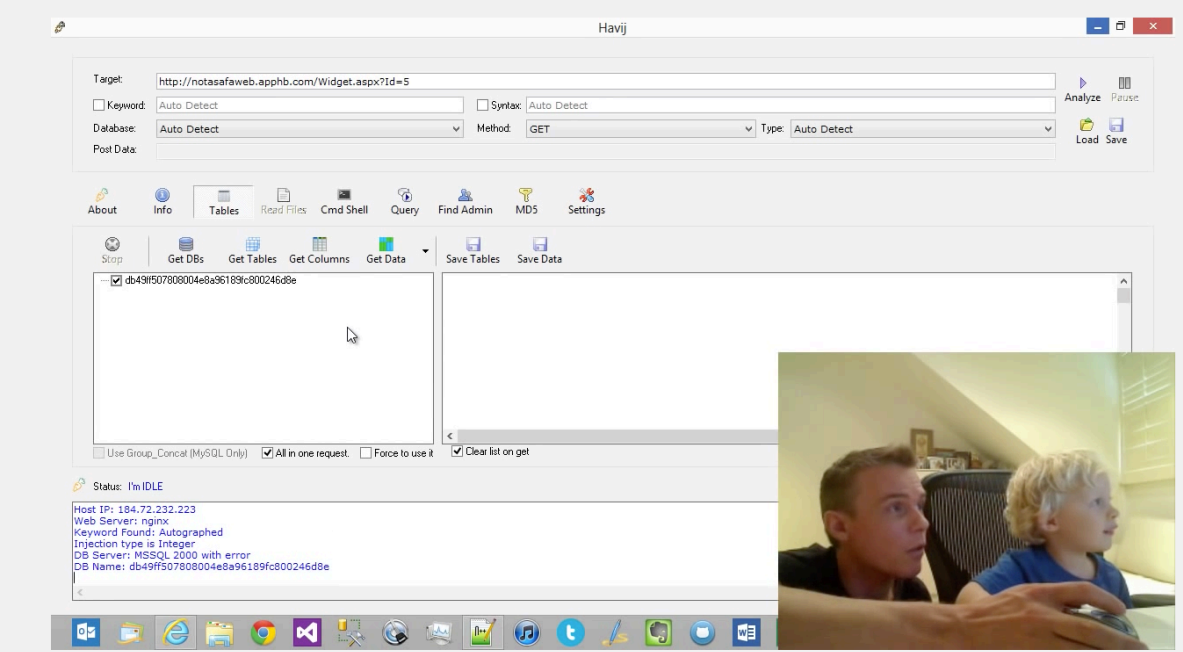
- 网络安全
- 应用安全
- 主机安全
- 业务安全
- 员工意识

印象深刻的两件事

- 女主播事件
- 钓鱼邮件事件

企业可能遇到的对手

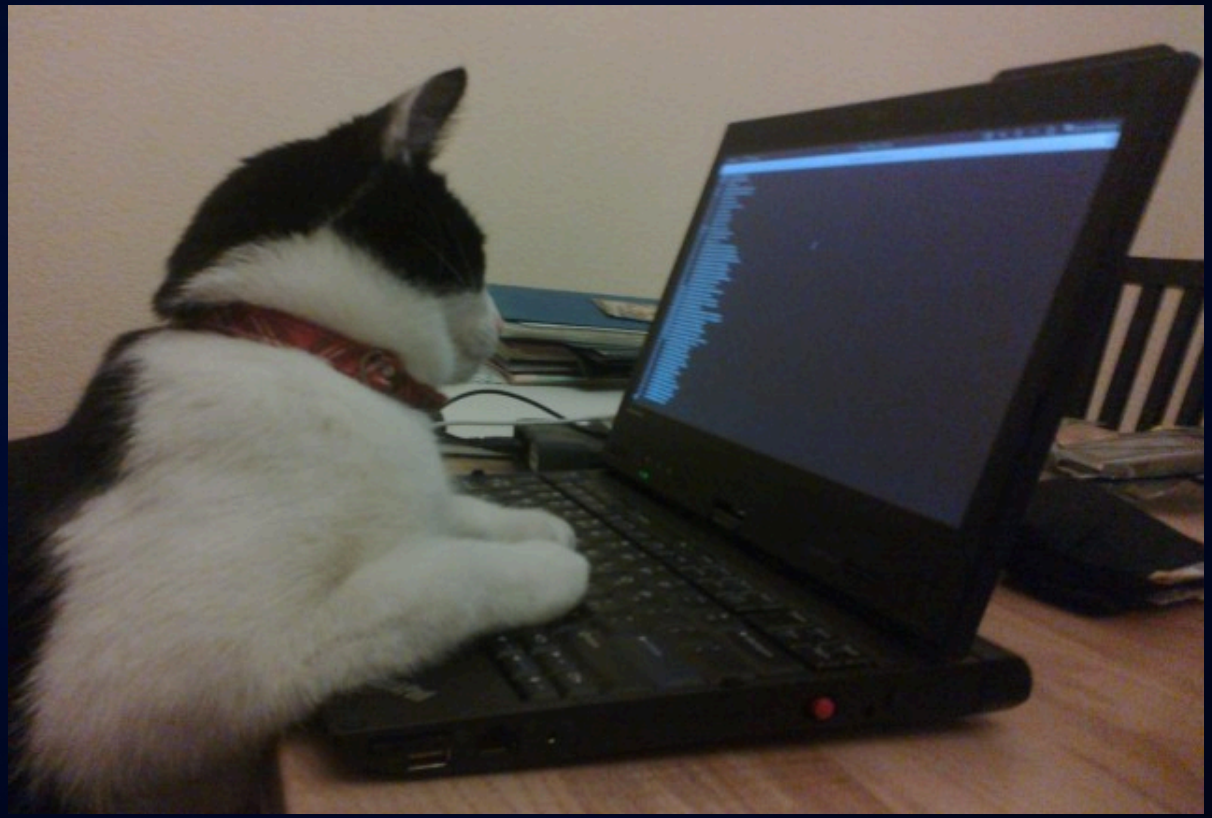
- 职业渗透测试人员
- 保持正义的白帽子
- 某种目的的黑客
- 也可能是个孩子



Hacking is child's play - SQL injection with Havij by 3 year old

企业可能遇到的对手

- 甚至是.....



攻防对抗的思考

- 攻击低龄化，防守依然老龄化？
- “攻”是一个点，“防”是一个面
- 防御了WEB攻击，那业务安全呢？
- WAF的防御面可以覆盖的更广吗？

传统WAF的特点

- 基于正则表达式引擎防御攻击
  - 容易误拦截漏拦截
  - 无法防御0day攻击
- 无法防御业务相关的安全问题
  - 薅羊毛
  - 越权漏洞
  - 防爬虫
  - 密码重置
  - 零元购

正则表达式引擎

- 正则表达式在Noam Chomosky里归于Type 3，属于表达最弱的语言

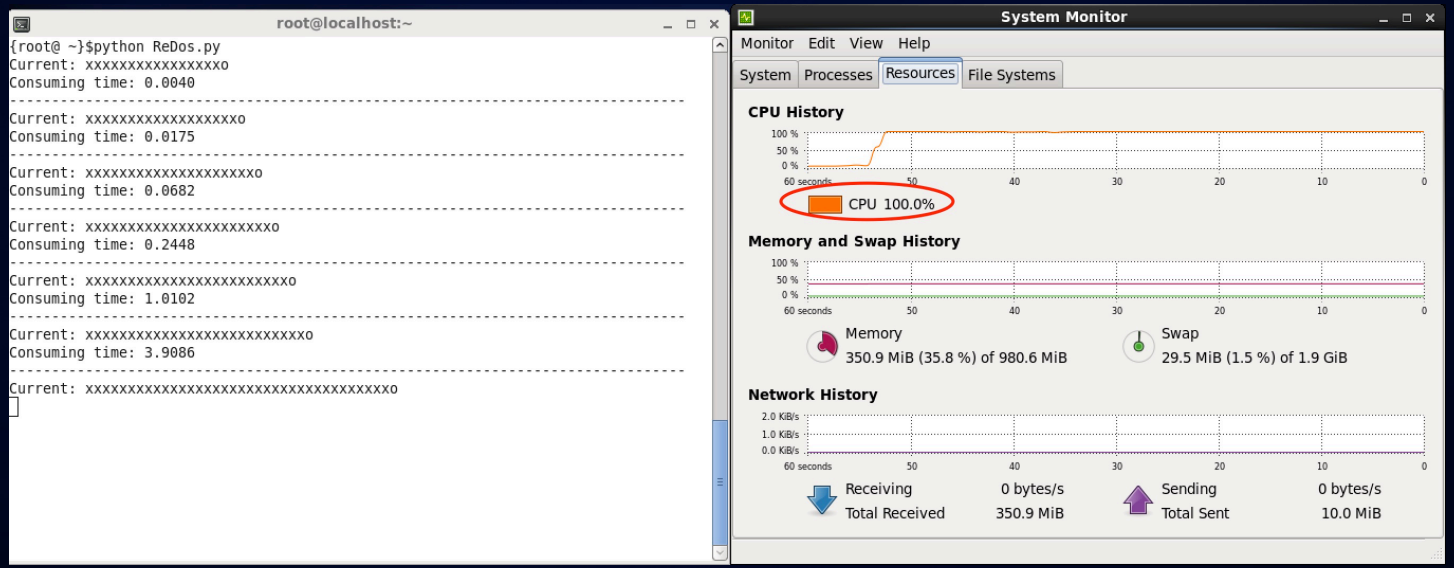
Grammar Type	Grammar Accepted	Language Accepted	Automaton
Type 0	Unrestricted grammar	Recursively enumerable language	Turing Machine
Type 1	Context-sensitive grammar	Context-sensitive language	Linear-bounded automaton
Type 2	Context-free grammar	Context-free language	Pushdown automaton
Type 3	Regular grammar	Regular language	Finite state automaton

正则表达式引擎

- 正则表达式引擎主要分为：确定型有穷自动机(DFA)，非确定型有穷自动机(NFA)。
- NFA反复吞吐字符串，速度慢，但支持backtracking
- DFA对文本里的每个字符只扫描一次，速度快

NFA （non-deterministic finite state machine）

- 正则表达式处理属于CPU密集型的操作
- ReDos（Regular expression Denial of Service）



DFA（deterministic finite state machine）

- 避免了ReDos的问题，但由于正则表达式描述性差，规则很难维护
- 防御的滞后性，出现漏洞才去更新规则，面对0day漏洞比较被动

```
{root@ ~}$  who''ami
root
{root@ ~}$  who""ami
root
{root@ ~}$  who``ami
root
{root@ ~}$  who````""``ami
root
{root@ ~}$  who''''''""``am''''i""
root
{root@ ~}$  who$()am$()i
root
{root@ ~}$  who$(echo "")ami
root
{root@ ~}$  who`echo ""`ami
root
{root@ ~}$  who`echo ""`am$(echo "")i
root
{root@ ~}$  ``w```echo""````h```echo""````o```echo""````a```echo ""````m```echo""````i``
root
```

正则引擎的误报和漏报直接影响到业务

- 误报会影响业务的正常流量
- 漏报会降低对防御的预期

而面对业务安全又表现的束手无策

- 😭

新一代WAF的特性

- 抛弃传统规则
- 关注业务安全
- 增加溯源能力

使用语义分析引擎

- 避免正则引擎误拦截导致的业务正常流量受影响
- 在一定程度上弥补正则引擎检测不到的攻击请求

希望拦截的请求

正则匹配了一条SQL注入特征的请求

Expression

```
/union[\s\S]+[\s\S]+select[\s\S]/g
```

Text

```
/*!union*/+/*!select*/+1,2,3—
```

不希望拦截的请求

正则也匹配了正常业务的请求

Expression

```
/union[\s\S]+[\s\S]+select[\s\S]/g
```

Text

```
in our union we should select the best worker
```

显然语义分析引擎可以很好的解决这样问题

```
~  python sqli_check.py request.txt
##################################################
>> sqli with fingerprint of 'X'
[+] /*!union*/+/*!select*/+1,2,3-
>> not sqli
[+] in our union we should select the best worker
##################################################
```

使用机器学习

- 采用有监督学习(supervised learning)方式，请求参数抽取数据->分词统计->通过黑白样->参数量化->二分类
- 训练样本保持正态分布，从而避免过拟合

序号	param
1	_%3D1498591621808
2	code%3Dzs_000001%2Czs_399001%2Czs_399006%26cb%3Dfort...
3	_%3D1498591951848%26list%3Dml_sh600030
4	6053%26ri%3Dzb6-00f%7E-04gUry-01h-0RC%26tn%3D1%26en...
5	b1498592370545%3D1
6	v%3D13111002
7	COLLCC%3D3442798258%26
8	t%3Dcheck%26rec%3Dstratus%26etyp%3Dconnect%26zone%3D...
9	cn_600022%2Ccn_600516%2Ccn_000002%2Ccn_600519%2Ccn_...
10	_%3D1498179095094%26list%3Dsh600030

序号	param
1	Search%3D%3C/script%3E%3Cimg/%2A%00/src%3D%22worksi...
2	symbol%3D%3Ch1%3E%3Cscript%3Ealert%28/hacked/%29%3C...
3	query%3D%3CIMG%2B%22%22%22%3E%3CSCRIPT%3Ealert...
4	ReturnUrl%3Dh[redacted]fr/recherche/recherche-globale/...
5	_lang%3D%22%3E%3Cscript%3Ealert%28document.cookie%29...
6	language%3D%22%3E%3C/script%3E%22%3E%27%3E%3Cscri...
7	q%3Dbentley%26stylesheet%3D%22%3E%3Cscript%3Ealert%28...
8	option%3Dcom_wdshop%26view%3Duserinfo%26ajax_json%3Daj...
9	CT_ORIG_URL%3D/arena/%22%3E%3Cscript%3Ealert%281337...
10	query%3DSearch...%26Product%3D%27%22--%3E%3C/style%3...

数据风控

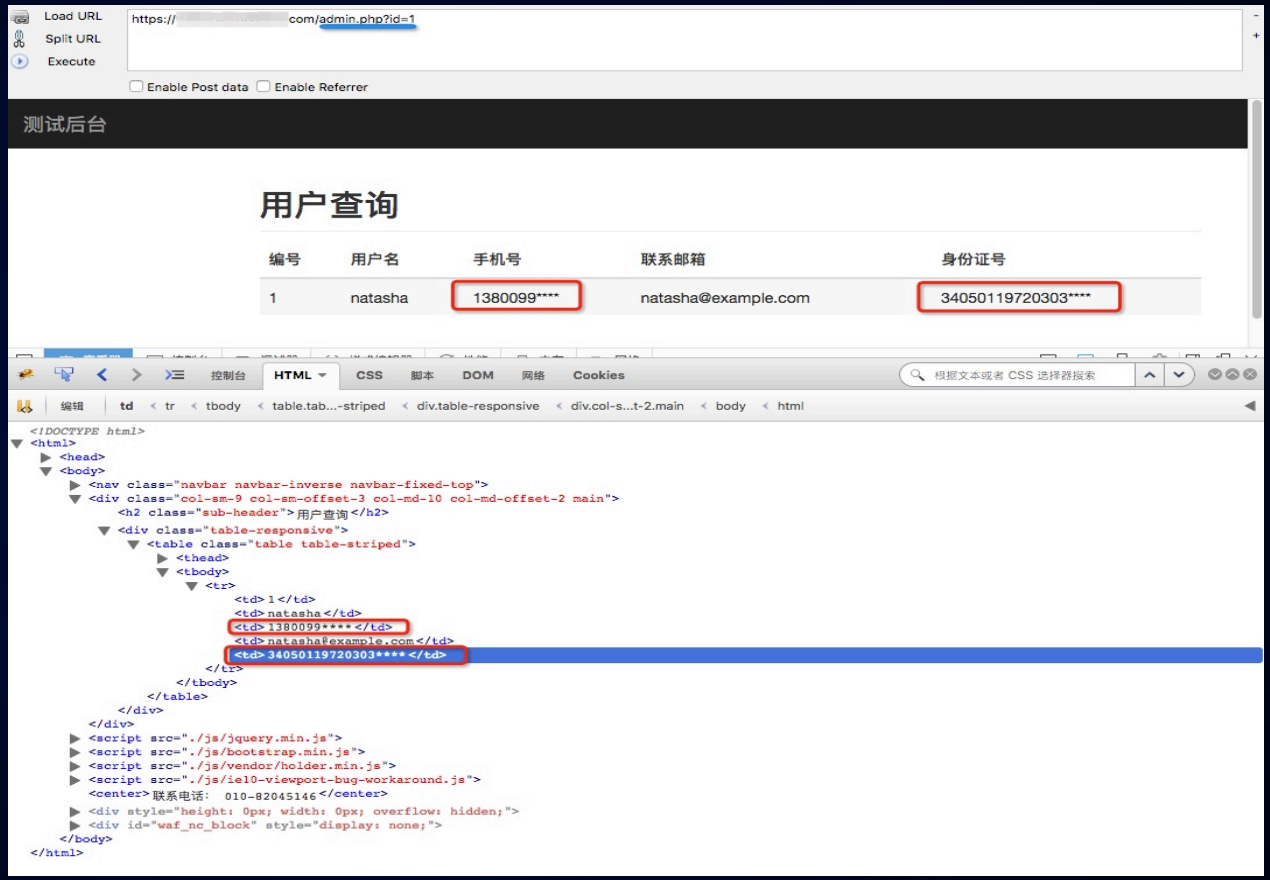
- 垃圾注册
- 短信接口滥用
- 扫库、撞库
- 刷票、恶意投票
- 薅羊毛

数据风控

- 信誉库：IP、手机号、设备
- 行为识别的算法：请求上下文分析
- 人机识别算法：JSSDK、滑块等

防信息泄漏

- 越权查看漏洞
- 未授权访问



防爬虫

- 反防爬传统手段

  - 非固定频率
  - 随机UA
  - 分散IP

- 反防爬对抗上升

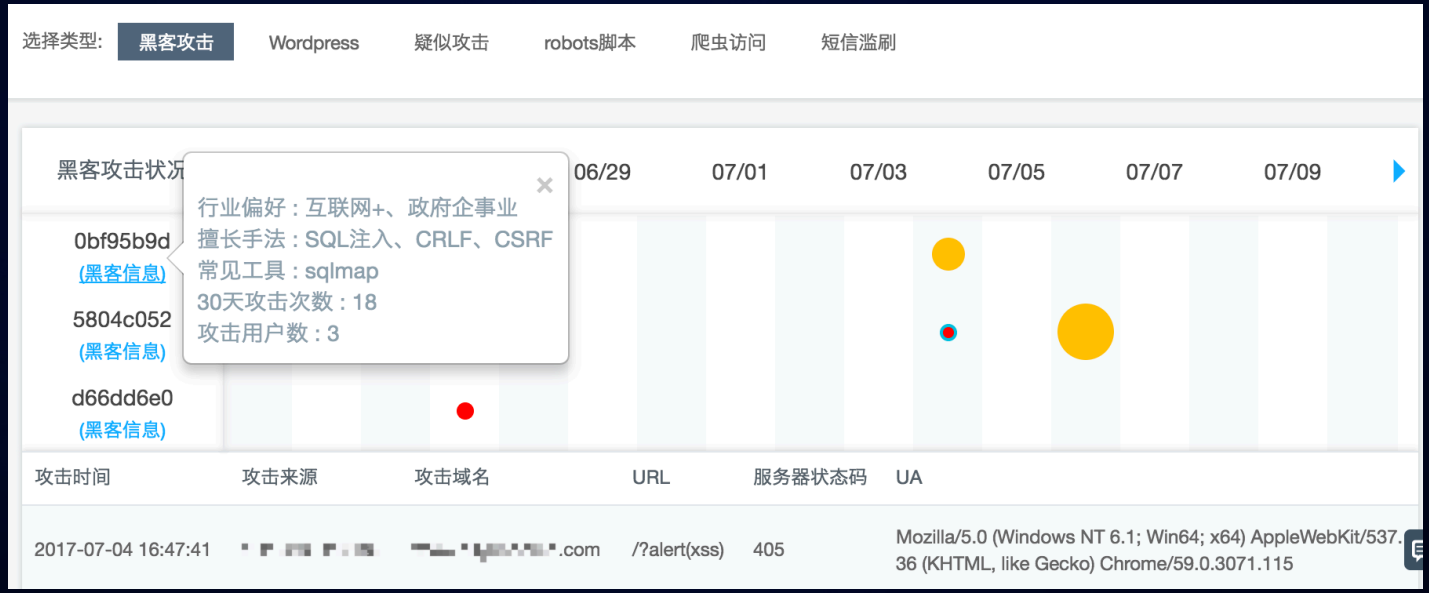
  - 低频
  - 慢速
  - 多源

防爬虫

- 特征库：IP、UA、手机号等
- 人机识别算法：结合业务风控、JSSDK、滑块等
- 爬虫行为分析：周期性、孤立会话等

黑客画像

- 通过规则建立模型，利用大数据关联分析
- 描绘出攻击者行为，为溯源提供依据



随着攻防对抗水平的不断提高，我们可能面对更对未知的风险。

在这种环境下，业务安全会面临更大的风险和挑战，在这种新趋势下，

WAF可以为企业安全承担更大的责任。

Thank you