

# Machine Learning in Cyber Security

## 利用机器学习，捍卫网络安全

**张佳彦**

趋势科技 数据科学家



**01**

网络威胁的演化与机器学习

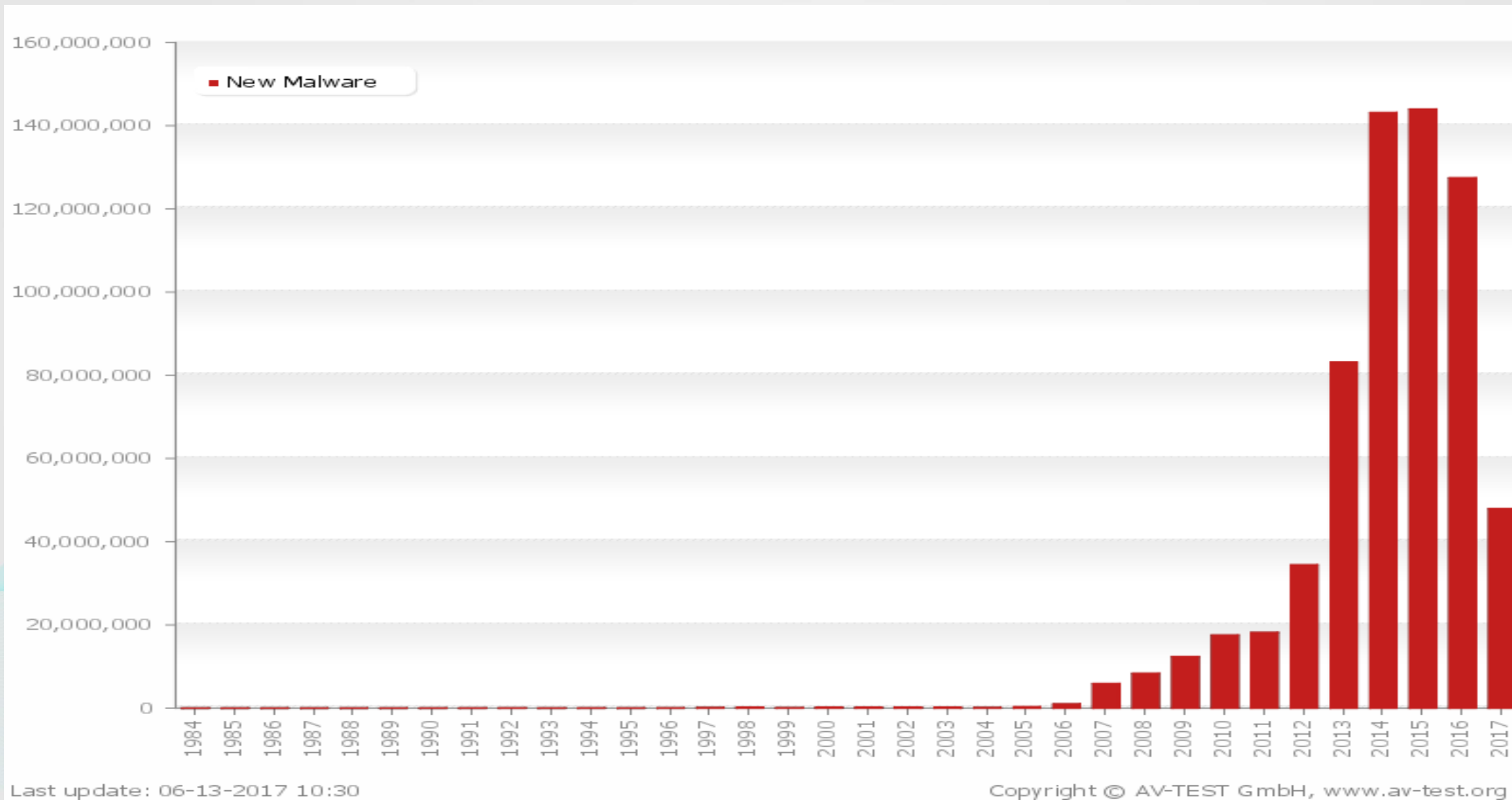
应用机器学习，防卫网络安全

**02**

**03**

机器学习是万灵丹？

# 网络威胁的演化





# 网络威胁的演化与机器学习

时代	形态	目的	数量	实例	主要解决方案
初期 ~2006	爆发型	成名	少	Melissa	Signature
中期 2006~2012	潜伏型	资源 资讯	多	Botnet Stuxnet	1-N

# 网络威胁的演化与机器学习

时代	形态	目的	数量	实例	主要解决方案
初期 ~2006	爆发型	成名	少	Melissa	Signature
中期 2006~2012	潜伏型	资源 资讯	多	Botnet Stuxnet	1-N
现代 2012~	勒索软件	金钱	极多	WannaCry	ML

经济学家亨利·乔治（ Henry George ）曾经说过这样的话：

人要吃小鸡，鹰也要吃小鸡

**鹰多吃**一只小鸡世界上的小鸡就**少**一只

**人多吃**一只小鸡世界上的小鸡就会**多**一只！



# 应用机器学习 防卫网络安全



巨量资料  
基础建设

网安专家  
知识

机器学习  
专家知识

高传真  
机器学习

数据

特征

演算法

模型

Ransom-Tescrypt  
Size: **326144** bytes

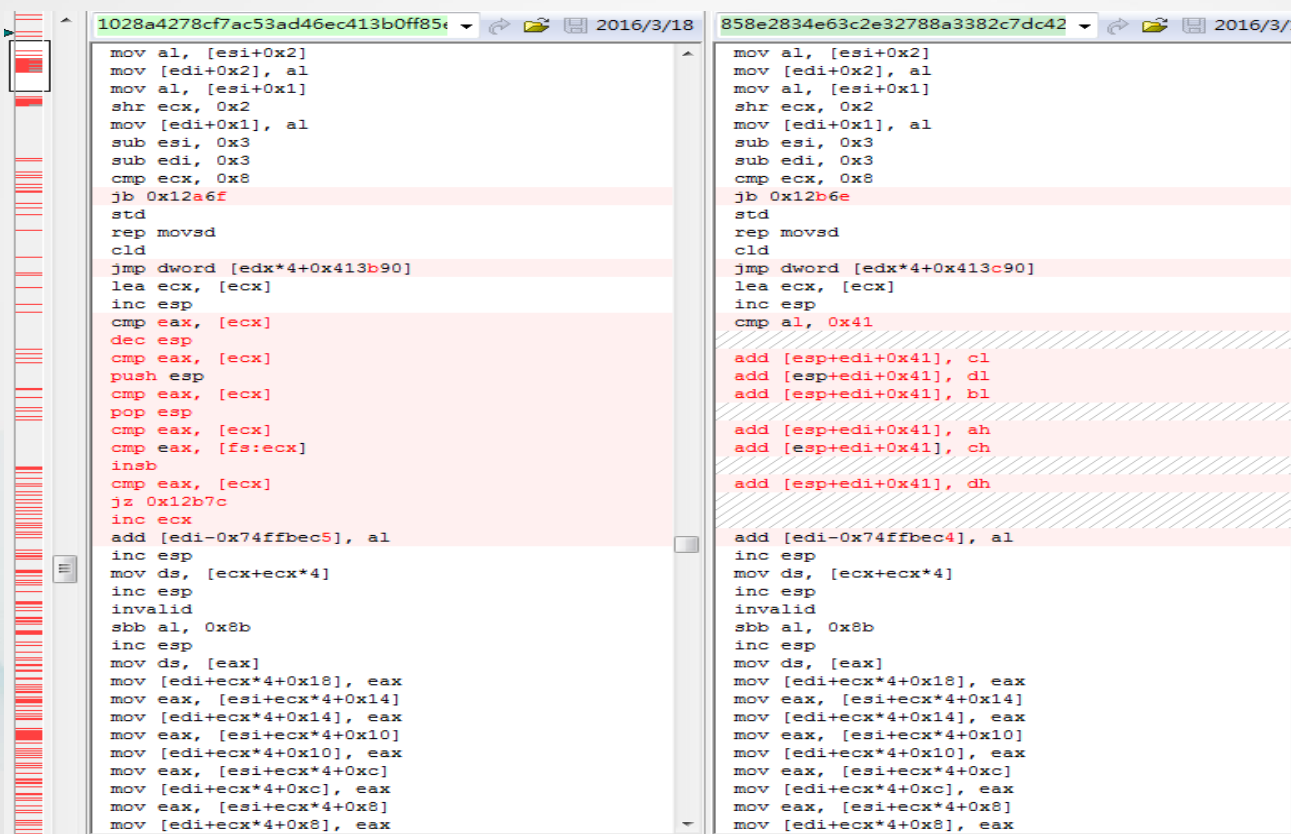
Ransom-Tescrypt.H  
Size: **196380** bytes

0004ba0	158b	d230	0041	5589	a1b0	d234	0041	4589	0004ba0	0d8b	e1c4	0041	8d89	ff4c	ffff	158b	e1c8
0004bb0	8bb4	380d	41d2	8900	b84d	158b	d23c	0041	0004bb0	0041	9589	ff50	ffff	cca1	41e1	8900	5485
0004bc0	5589	a1bc	d240	0041	4589	8bc0	440d	41d2	0004bc0	ffff	8bff	d00d	41e1	8900	588d	ffff	8bff
0004bd0	8900	c44d	158b	d248	0041	5589	a1c8	d24c	0004bd0	d415	41e1	8900	5c95	ffff	a1ff	e1d8	0041
0004be0	0041	4589	8bcc	500d	41d2	8900	d04d	158b	0004be0	8589	ff60	ffff	0d8b	e1dc	0041	8d89	ff64
0004bf0	d254	0041	5589	a1d4	d258	0041	4589	8bd8	0004bf0	ffff	158b	e1e0	0041	9589	ff68	ffff	e4a1
0004c00	5c0d	41d2	8900	dc4d	158b	d260	0041	5589	0004c00	41e1	8900	6c85	ffff	8bff	e80d	41e1	8900
0004c10	a1e0	d264	0041	4589	8be4	680d	41d2	8900	0004c10	708d	ffff	8bff	ec15	41e1	8900	7495	ffff
0004c20	e84d	158b	d26c	0041	5589	a1ec	d270	0041	0004c20	a1ff	e1f8	0041	8589	ff78	ffff	0d8b	e1fc
0004c30	4589	8bf0	740d	41d2	8900	f44d	158b	d198	0004c30	0041	8d89	ff7c	ffff	158b	e200	0041	5589
0004c40	0041	5589	a1f8	d04c	0041	8589	fed4	ffff	0004c40	a180	e204	0041	4589	8b84	080d	41e2	8900
0004c50	0d8b	d048	0041	8d89	fed8	ffff	158b	d044	0004c50	884d	158b	e20c	0041	5589	a18c	e210	0041
0004c60	0041	9589	fedc	ffff	40a1	41d0	8900	e085	0004c60	4589	8b90	140d	41e2	8900	944d	158b	e218
0004c70	fffe	8bff	3c0d	41d0	8900	e48d	fffe	8bff	0004c70	0041	5589	a198	e21c	0041	4589	8b9c	200d
0004c80	3815	41d0	8900	e895	fffe	a1ff	d034	0041	0004c80	41e2	8900	a04d	158b	e224	0041	5589	a1a4
0004c90	8589	feec	ffff	0d8b	d030	0041	8d89	fef0	0004c90	e228	0041	4589	8ba8	2c0d	41e2	8900	ac4d
0004ca0	ffff	158b	d050	0041	9589	fef4	ffff	00a1	0004ca0	158b	e230	0041	5589	a1b0	e234	0041	4589
0004cb0	41d0	8900	a885	fffe	8bff	040d	41d0	8900	0004cb0	8bb4	380d	41e2	8900	b84d	158b	e23c	0041
0004cc0	ac8d	fffe	8bff	0815	41d0	8900	b095	fffe	0004cc0	5589	a1bc	e240	0041	4589	8bc0	440d	41e2
0004cd0	a1ff	d00c	0041	8589	feb4	ffff	0d8b	d010	0004cd0	8900	c44d	158b	e248	0041	5589	a1c8	e24c
0004ce0	0041	8d89	feb8	ffff	158b	d014	0041	9589	0004ce0	0041	4589	8bcc	500d	41e2	8900	d04d	158b
0004cf0	febc	ffff	18a1	41d0	8900	c085	fffe	8bff	0004cf0	e254	0041	5589	a1d4	e258	0041	4589	8bd8
0004d00	1c0d	41d0	8900	c48d	fffe	8bff	2015	41d0	0004d00	5c0d	41e2	8900	dc4d	158b	e260	0041	5589
0004d10	8900	c895	fffe	a1ff	d024	0041	8589	fecc	0004d10	a1e0	e264	0041	4589	8be4	680d	41e2	8900
0004d20	ffff	0d8b	d028	0041	8d89	fed0	ffff	158b	0004d20	e84d	158b	e26c	0041	5589	a1ec	e270	0041
0004d30	d284	0041	9589	fdd8	ffff	88a1	41d2	8900	0004d30	4589	8bf0	740d	41e2	8900	f44d	158b	e198
0004d40	dc85	fffd	8bff	8c0d	41d2	8900	e08d	ffff	0004d40	0041	5589	a1f8	e04c	0041	8589	fed4	ffff
0004d50	8bff	9015	41d2	8900	e495	fffd	a1ff	d294	0004d50	0d8b	e048	0041	8d89	fed8	ffff	158b	e044

16进制码  
元数据

Ransom-Tescrypt  
Size: **326144** bytes

Ransom-Tescrypt.H  
Size: **196380** bytes

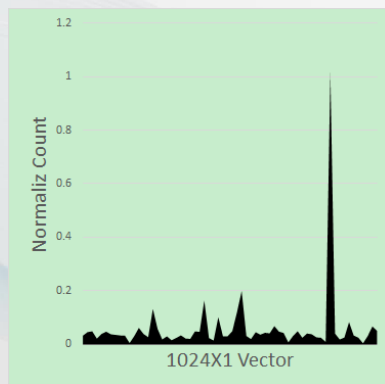
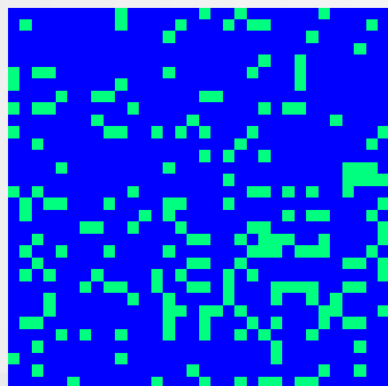


```
1028a4278cf7ac53ad46ec413b0ff85e 2016/3/18
mov al, [esi+0x2]
mov [edi+0x2], al
mov al, [esi+0x1]
shr ecx, 0x2
mov [edi+0x1], al
sub esi, 0x3
sub edi, 0x3
cmp ecx, 0x8
jb 0x12a6f
std
rep movsd
cld
jmp dword [edx*4+0x413b90]
lea ecx, [ecx]
inc esp
cmp eax, [ecx]
dec esp
cmp eax, [ecx]
push esp
cmp eax, [ecx]
pop esp
cmp eax, [ecx]
cmp eax, [fs:ecx]
insb
cmp eax, [ecx]
jz 0x12b7c
inc ecx
add [edi-0x74ffbec5], al
inc esp
mov ds, [ecx+ecx*4]
inc esp
invalid
sbb al, 0x8b
inc esp
mov ds, [eax]
mov [edi+ecx*4+0x18], eax
mov eax, [esi+ecx*4+0x14]
mov [edi+ecx*4+0x14], eax
mov eax, [esi+ecx*4+0x10]
mov [edi+ecx*4+0x10], eax
mov eax, [esi+ecx*4+0xc]
mov [edi+ecx*4+0xc], eax
mov eax, [esi+ecx*4+0x8]
mov [edi+ecx*4+0x8], eax

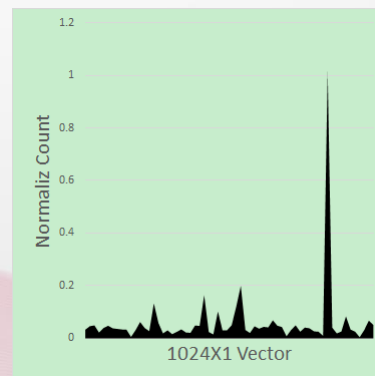
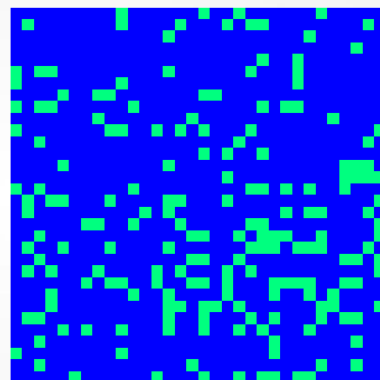
858e2834e63c2e32788a3382c7dc42 2016/3/18
mov al, [esi+0x2]
mov [edi+0x2], al
mov al, [esi+0x1]
shr ecx, 0x2
mov [edi+0x1], al
sub esi, 0x3
sub edi, 0x3
cmp ecx, 0x8
jb 0x12b6e
std
rep movsd
cld
jmp dword [edx*4+0x413c90]
lea ecx, [ecx]
inc esp
cmp al, 0x41
add [esp+edi+0x41], cl
add [esp+edi+0x41], dl
add [esp+edi+0x41], bl
add [esp+edi+0x41], ah
add [esp+edi+0x41], ch
add [esp+edi+0x41], dh
add [edi-0x74ffbec4], al
inc esp
mov ds, [ecx+ecx*4]
inc esp
invalid
sbb al, 0x8b
inc esp
mov ds, [eax]
mov [edi+ecx*4+0x18], eax
mov eax, [esi+ecx*4+0x14]
mov [edi+ecx*4+0x14], eax
mov eax, [esi+ecx*4+0x10]
mov [edi+ecx*4+0x10], eax
mov eax, [esi+ecx*4+0xc]
mov [edi+ecx*4+0xc], eax
mov eax, [esi+ecx*4+0x8]
mov [edi+ecx*4+0x8], eax
```

操作码

Ransom-Tescrypt  
Size: 326144 bytes



Ransom-Tescrypt.H  
Size: 196380 bytes



导入表

操作码  
统计

巨量资料  
基础建设

网安专家  
知识

机器学习  
专家知识

高传真  
机器学习

数据

特征

演算法

模型

巨量资料  
基础建设

机器学习  
专家知识

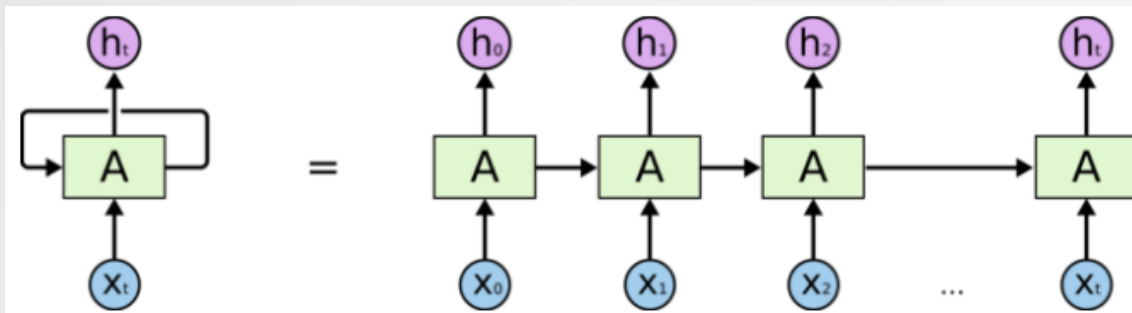
高传真  
机器学习

数据

演算法

模型

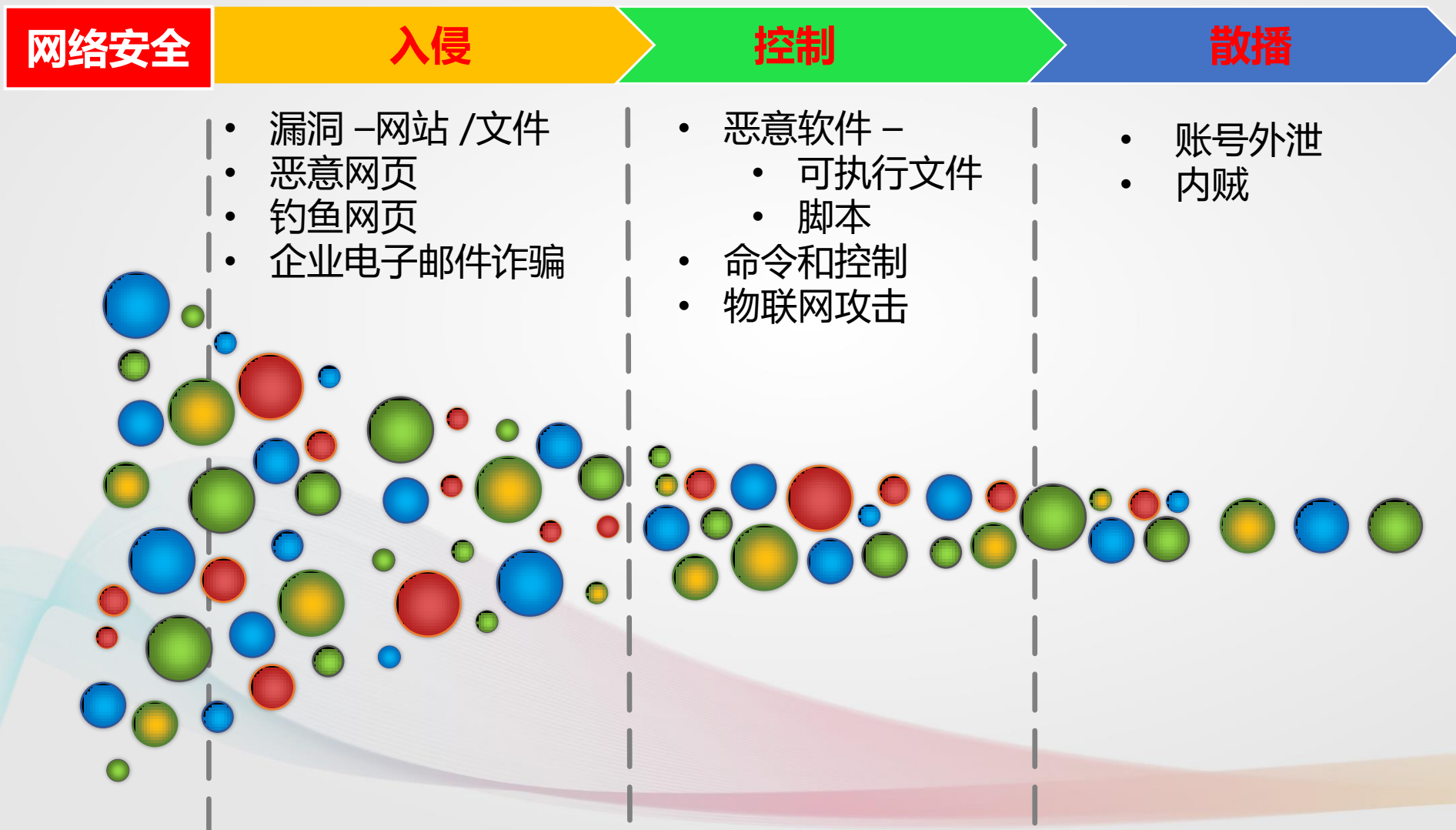
- 使用深度学习LSTM侦测恶意网址



```
htxp://www.tma.tw:80/ltk/106600411.pdf  
00000000000011111000000000111111111111
```

```
htxps://cdn.fbsbx.com:443/v/t59.2708-21/.../ch14-Fluid.exe  
0000000000010000000000000000001111111111...1111111111111111
```

# 网络安全攻击层级





# 利用机器学习，防卫网络安全



支持向量机



支持向量机



提升树



支持向量机



深度学习



漏洞



文件



邮件欺诈



钓鱼



恶意链接

- 漏洞模型
- 包含 Neutrino, Rig, Sundown, Kaixin, Hunter,等
- 执行于100G吞吐量

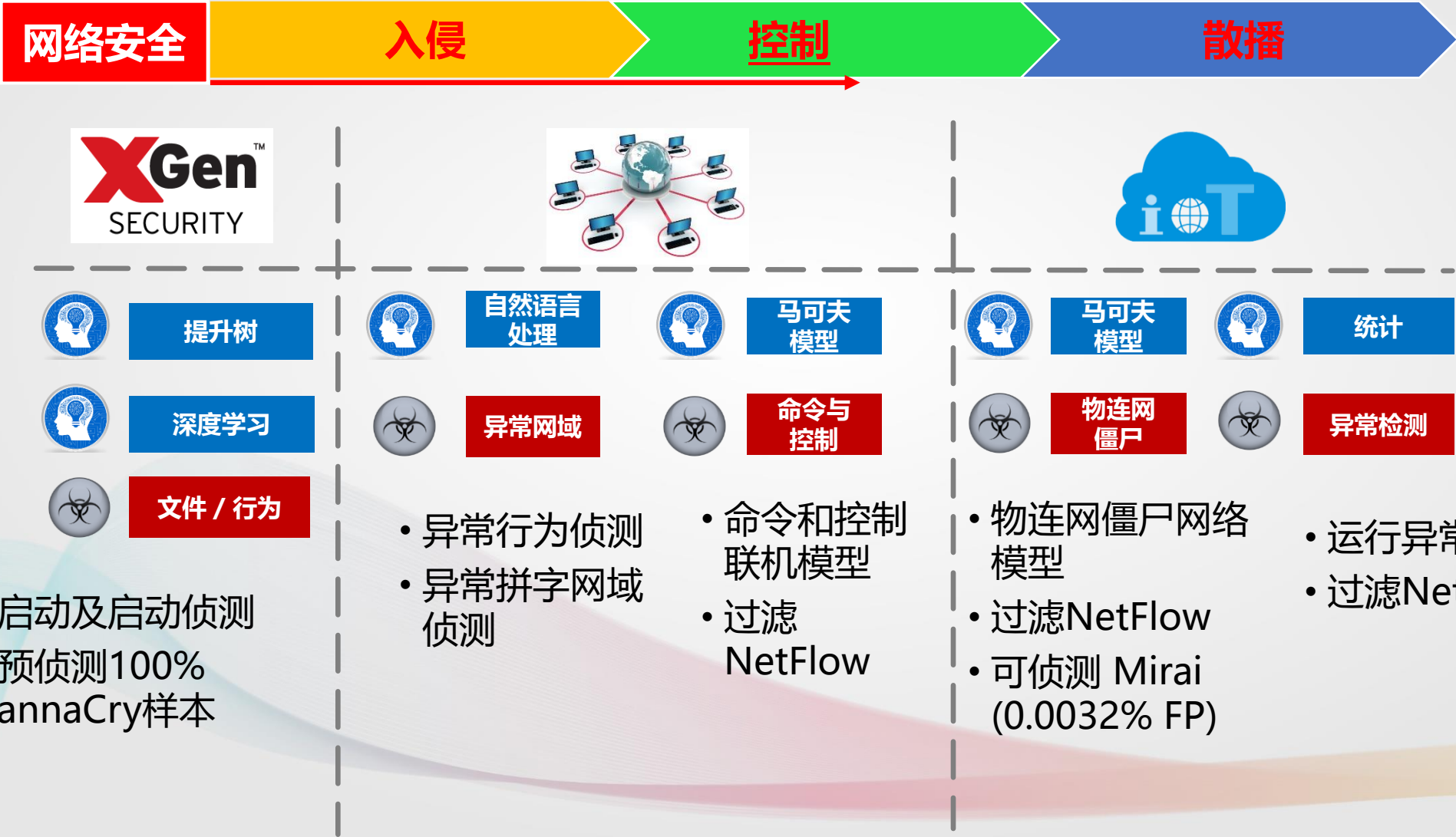
- 宏模型
- 包含社交工程攻击,下载器等.

- VIP写作格式模型
- 包含CEO诈骗,付款劫持等.

- 保护394个知名网站
- 每天侦测 15 万个钓鱼网页

- 每天70亿次用户拜访
- 每天阻挡450 万次恶意网页攻击

# 利用机器学习，防卫网络安全



# 利用机器学习，防卫网络安全

网络安全

入侵

控制

散播



置信度传播

+

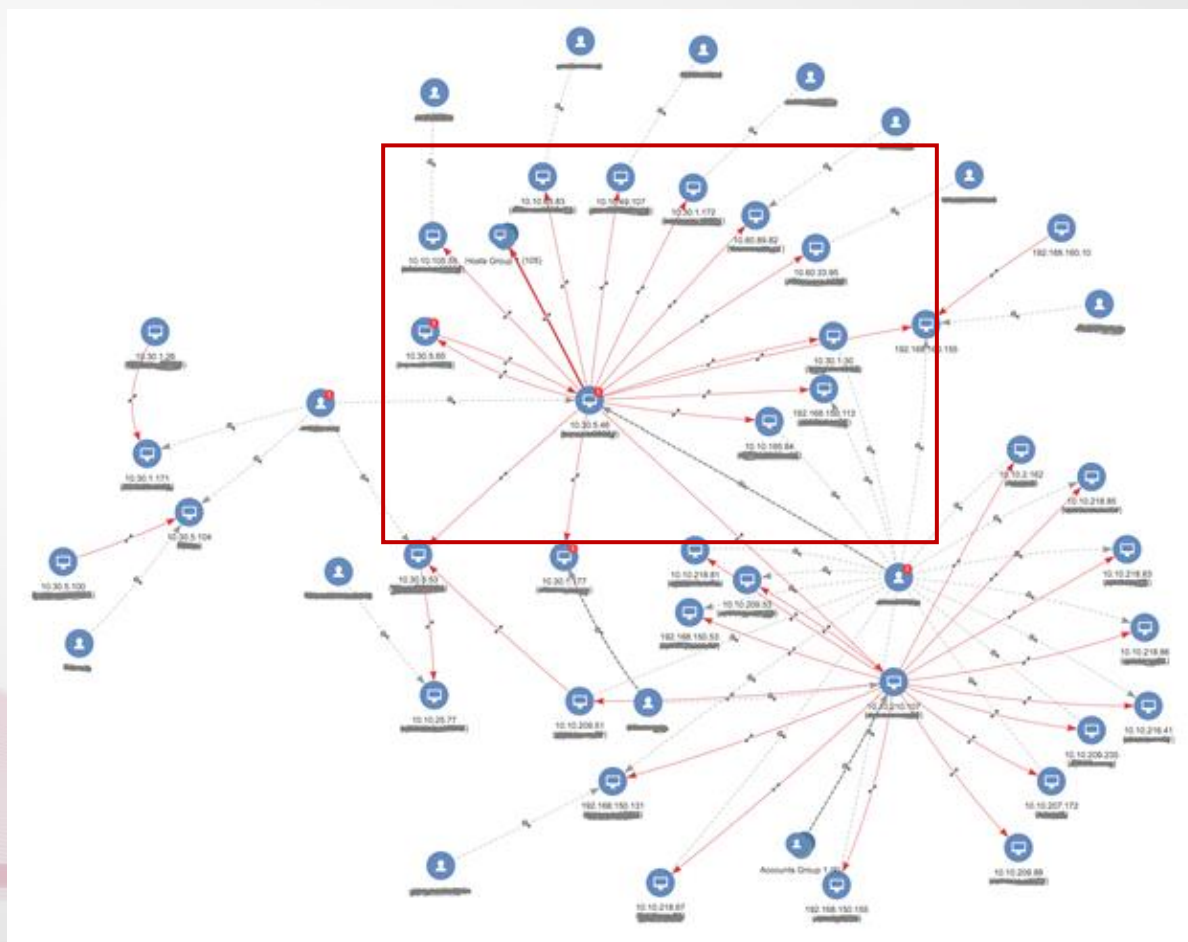
马尔可夫链



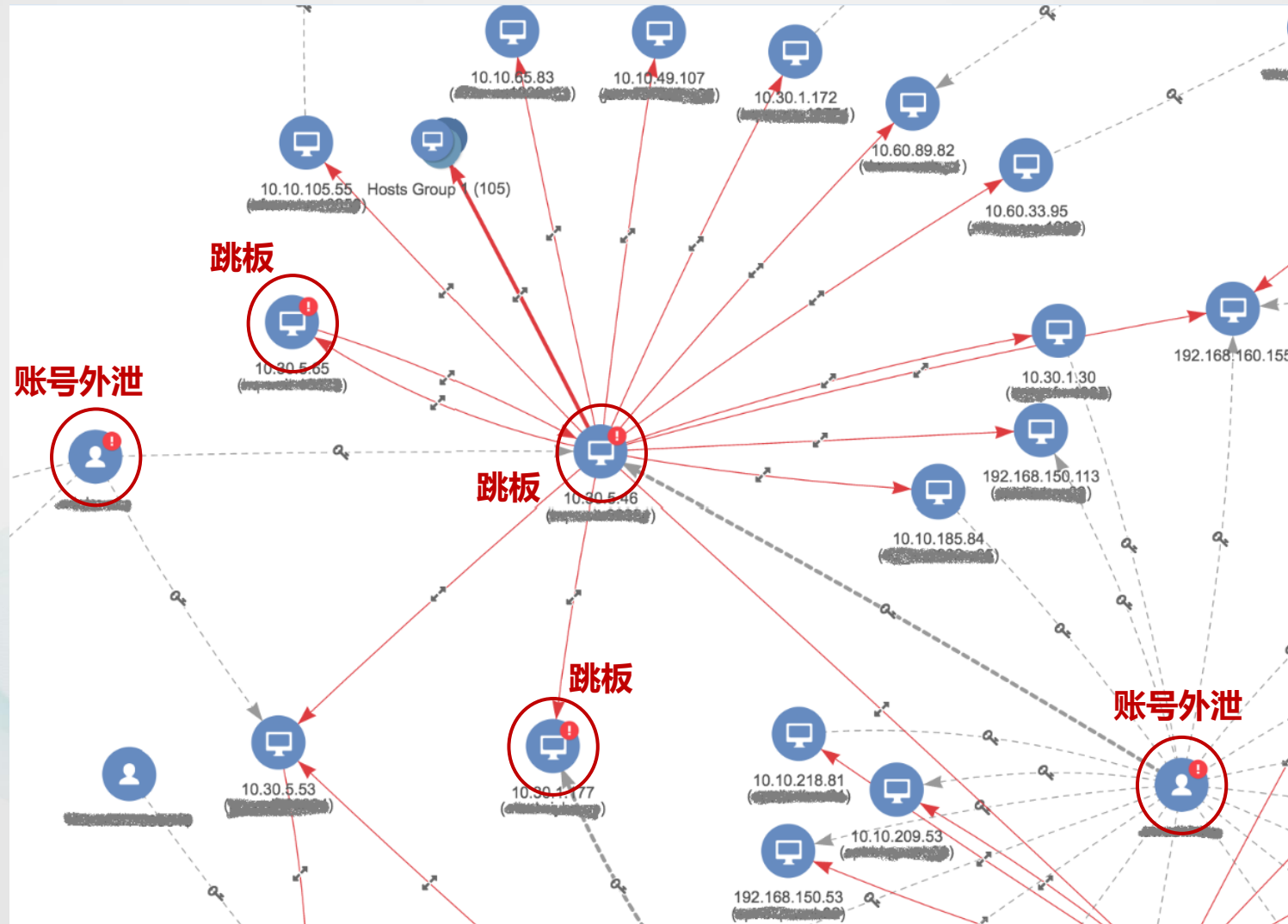
账号外泄

主机入侵

- 强调严重事件，账号外泄，主机入侵
- 提供网络狙杀炼信息



# 实例



# 机器学习是万灵丹？

# 机器学习是万灵丹?





- 问题: 恶意软件与良性软件总是有正确答案吗?
- 一个软件会下载一个档案并执行
  - 恶意软件?
    - 勒索软体下载器
  - 良性软件?
    - 网路会议安装软件

# 误判与机器学习

- 传统侦测误判率
  - 1/10,000,00
- 机器学习误判率
  - 1/1,000 ~ 1/10,000
- 扫描计算机里的所有档案?
  - 1只恶意软件VS 20,000个良性软件 → 1正判/2~20误判
- 扫描所有的下载文件?
  - 1只恶意软件VS 100个良性软件 → 1正判/0.1~0.01误判

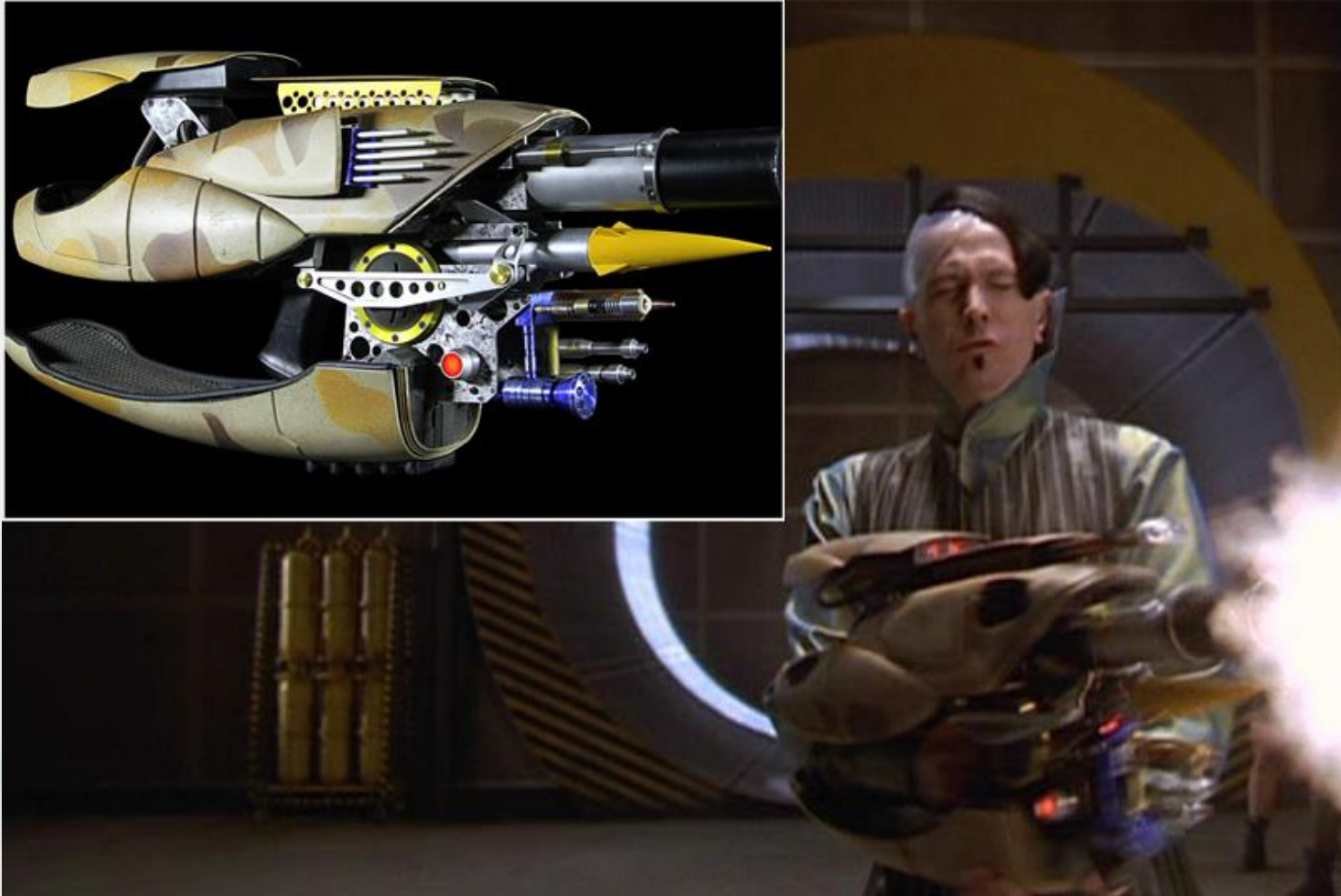




## 降噪

- 在正确的地方使用机器学习
  - 在地白名单过滤
  - 通道 – 网络下载/邮件附件/可疑行为
  - 普查 – 新出现文件
  - 签名 – 无信任签名

# 机器学习是万灵丹?



# Thank You



# C3