S&P-2018

Understanding Linux Malware

S&P-2017

A Lustrum of malware network communication: Evolution and insights

NDSS-2016

TLS in the wild: An Internet-wide analysis of TLS-based protocols for electronic communication

CCS-2016

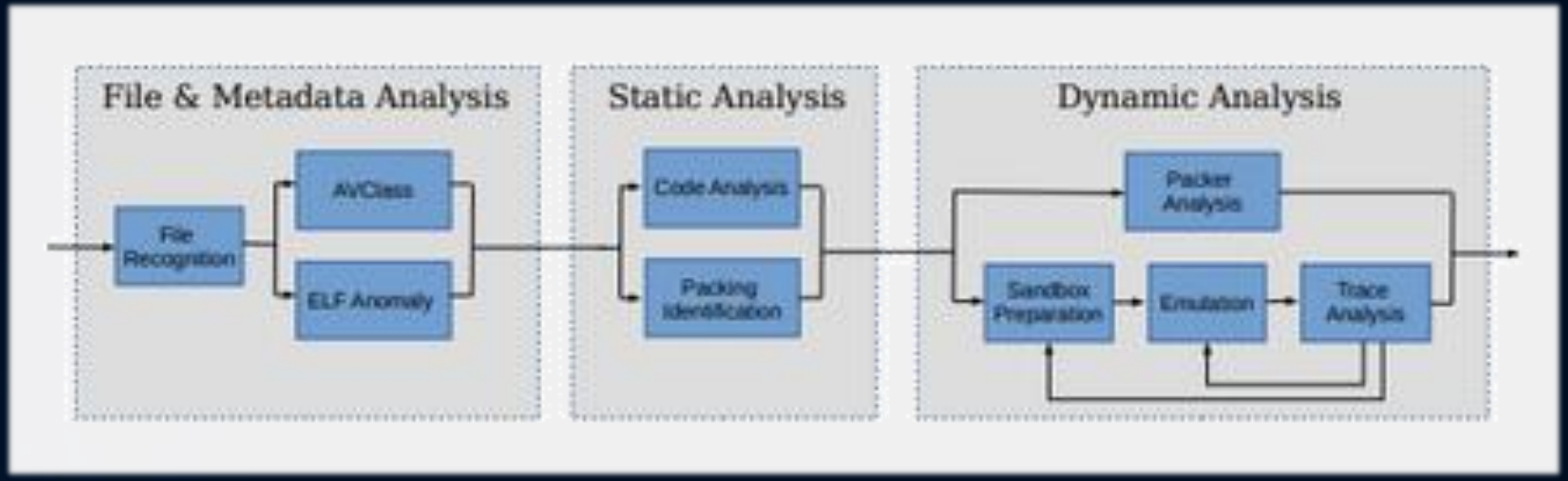
PhishEye: Live Monitoring of Sandboxed Phishing Kits

Understanding Linux Malware

深入理解 LINUX 恶意软件

存在的问题及挑战

- 没有针对 Linux 平台上恶意软件的大规模测量
- 安全社区在 2014 年才开始注意这些问题
- 体系结构复杂，数十种不同的体系结构
- ELF 规范允许程序任意指定 Loader
- 不同操作系统的针对性开发（OS/ABI）
- 静态链接难以分析识别，分析环境与权限问题



- 实现自定义 ELF 格式解析器
- 使用 AVClass 对恶意软件的 VT 报告标签进行归类
- 使用 IDA Pro 提取代码度量（函数数量、熵、覆盖率、cyclomatic 复杂度）
- 基于 KVM 和 QEMU 支持多种体系架构的沙盒
- 根据 systemtap 实现内核探针的插桩
- 基于 Unicorn 仿真多体系结构的指令解决壳识别与脱壳问题

体系结构与 ELF 文件解析

Architecture	Samples	Percentage
X86-64	3018	28.61%
MIPS I	2120	20.10%
PowerPC	1569	14.87%
Motorola 68000	1216	11.53%
Sparc	1170	11.09%
Intel 80386	720	6.83%
ARM 32-bit	555	5.26%
Hitachi SH	130	1.23%
AArch64 (ARM 64-bit)	47	0.45%
others	3	0.03%

- 83% 的恶意软件可以和某个家族关联起来
- 僵尸网络超过 69% 的样本
- 类别有后门、勒索、加密货币矿工、Rootkit、RAT 等

Technique	Samples	Percentage
Segment header table pointing beyond file data	1	0.01%
Overlapping ELF header/segment	2	0.02%
Wrong string table index (`e_shstrndx`)	60	0.57%
Section header table pointing beyond file data	178	1.69%
Total Corrupted	211	2.00%

- 面对异常文件与无效文件，只有 IDA Pro 可以正常解析 ELF 文件

持久性策略

- 21% 的恶意软件至少实现了一个持久性策略，往往使用多种技术实现目标
- 子系统初始化
- 基于时间的执行
- 文件感染与替换
- 用户文件更改

Path	Samples	
	w/o root	w/ root
/etc/rc.d/rc.local	-	1393
/etc/rc.conf	-	1236
/etc/init.d/	-	210
/etc/rcX.d/	-	212
/etc/rc.local	-	11
systemd service	-	2
~/.bashrc	19	8
~/.bash_profile	18	8
X desktop autostart	3	1
/etc/cron.hourly/	-	70
/etc/crontab	-	70
/etc/cron.daily/	-	26
crontab utility	6	6
File replacement	-	110
File infection	5	26
Total	1644 (21.10%)	

欺骗与权限问题

- 50% 的样本在内存中有着不同的名字，冒用良性程序的名字
- 其中 11% 使用了公用程序的名字
- 其中 88% 使用了空、虚构、随机的名字

Different behavior	Samples	Percentage
Execute privileged shell command	579	21.96%
Drop a file into a protected directory	426	16.15%
Achieve system-wide persistence	259	9.82%
Tamper with Sandbox	61	2.31%
Delete a protected file	47	1.78%
Run *ptrace* request on another process	10	0.38%

- 25% 的样本都会尝试不同权限执行
- 其中 89% 的样本两次执行的行为存在差异

权限提升

- CVE-2016-5195 是最常用的漏洞，总共有 52 个 ELF 程序试图在沙盒中利用它
- 还发现了五次利用 CVE-2015-1328 的尝试
- 使用 root 权限重新执行的 2637 个恶意软件示例中，只有 15 个成功加载了内核模块，并且没有一个执行卸载过程
- 所有这些情况涉及标准 ip_tables.ko

加壳与多态

- Linux 中只有少数被提出，大多数都是概念性证明，只有 UPX 例外
- Vanilla UPX 及其变体是数据集中最普遍的加壳形式，380 个加壳的二进制程序中，只有三个不属于这一类别

Process name	Samples	Percentage
Vanilla UPX	189	1.79%
Custom UPX Variant	188	1.78%
- Different Magic	129	
- Modified UPX strings	55	
- Inserted junk bytes	126	
- All of the previous	16	
Mumblehard Packer	3	0.03%

- 对壳的二次修改
- 自定义壳

进程交互

- 25% 的样本由一个进程组成，9% 的样本产生一个新进程，43% 的样本涉及三个进程(主要是用于创建守护进程时的“double-fork”模式)，其余 23% 的样本创建了更多的独立进程（多达 1684）

Shell command	Samples	Percentage
sh	400	5.13%
sed	243	3.12%
cp	223	2.86%
rm	216	2.77%
grep	214	2.75%
ps	131	1.68%
insmod	124	1.59%
chmod	113	1.45%
cat	93	1.19%
iptables	84	1.08%

- 13% 至少执行了一个外部 shell 命令
- 监视三种进程注入技术
- 监控多处重要配置文件位置

规避

Type of evasion	Samples	Percentage
Sandbox detection	19	0.24%
Processes enumeration *	259	3.32%
Anti-debugging	63	0.81%
Anti-execution	3	0.04%
Stalling code	0	-

Path	Detected Environments	#
/sys/class/dmi/id/product_name	VMware/VirtualBox	18
/sys/class/dmi/id/sys_vendor	QEMU	18
/proc/cpuinfo	CPU model/hypervisor flag	1
/proc/sysinfo	KVM	1
/proc/scsi/scsi	VMware/VirtualBox	1
/proc/vz and /proc/bc	OpenVZ container	1
/proc/xen/capabilities	XEN hypervisor	1
/proc/<PID>/mountinfo	chroot jail	1

- 沙盒检测
- 进程枚举
- 反调试
- 反执行
- 延迟执行

恶意软件多态性示例

- Tsunami 家族在这个数据集中有 743 个样本
- 为九种不同的架构编译
- 其中 86% 是静态链接的，13% 是剥离符号的
- 动态链接的样本依赖于不同的 loader
- 熵值从 1.85 到 7.99 不等
- 在熵较高的 19 个样品中，一个加壳 Vanilla UPX，其余使用相同算法的修改版本
- 只有 15% 样本测试了用户权限，或者得到了与特权相关的错误
- 17 个样本包含规避沙箱的代码，而其他所有样本都不包含规避功能

A Lustrum of malware network communication: Evolution and insights

恶意软件网络通信的进化与洞察

数据集构成

Dataset	Data	Count
Malware Executions	Samples with DNS	26.8 M
	FQDNs	11.5 M
	e2LDs	6.8 M
	IPs	1.4 M
VirusTotal	Reports	23.9 M
Passive DNS	Resource Records	5.2 B
	FQDNs	4.6 B
	e2LDs	2.9 M
	IPs	178.7 M
Public Blacklists	Distinct Blacklists	8
	e2LDs	320 K
Alexa	e2LDs	8 M
Expired Domains	e2LDs	179 M
DGArchive [12]	DGA FQDNs	50 M

- 五年（2011.1.1-2015.8.31）收集2680万样本的网络通信数据
- 两个商业和一个学术恶意软件源+50亿条DNS查询记录
- Passive DNS 数据源于美国一家大型ISP

黑名单数据与两类恶意程序

Blacklist	Target	Source
Abuse.ch	Malware, C&C.	[4]
Malware DL	Malware.	[11]
Blackhole DNS	Malware, Spyware.	[5]
sagadc	Malware, Fraud, SPAM.	[9]
hphosts	Malware, Fraud, Ad tracking.	[7]
SANS	Aggregate list.	[10]
itmate	Malicious Webpages.	[8]
driveby	Drive-by downloads.	[6]

- 过去七年一亿七千九百万域名到期日期
- 89%的样本在提交扫描的时候VT是已知的
- DGArchive可识别66个恶意软件家族

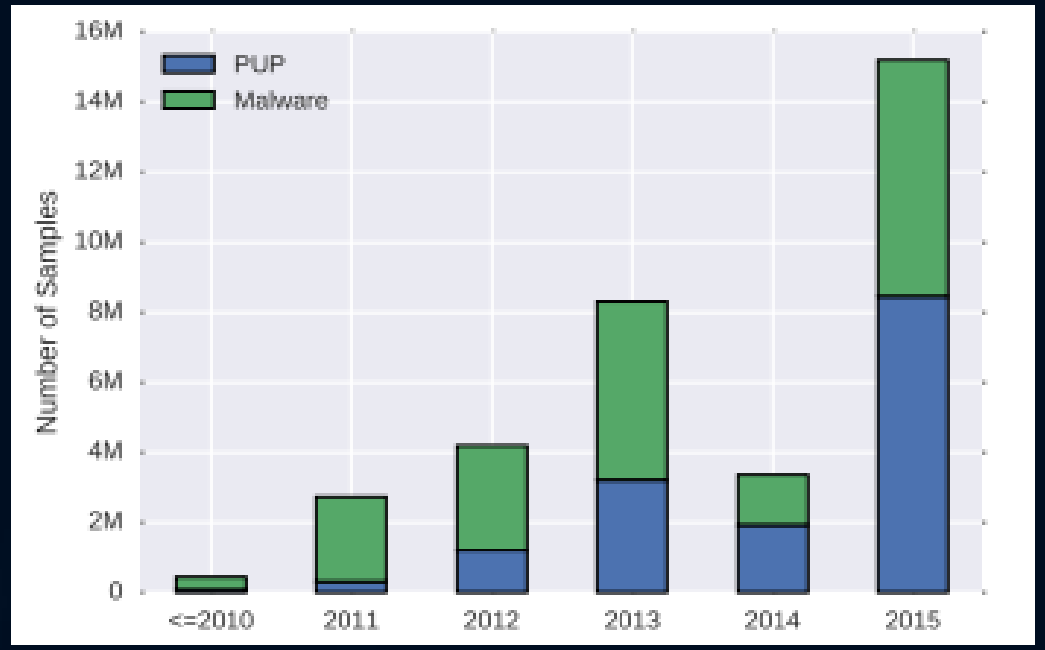
- Malware VS PUP(Potentially Unwanted Programs)
- PUP包括在浏览器中修改广告或搜索的广告软件、付费安装程序等

过滤与分类

- 删除没有任何反病毒厂商标记为恶意软件的 255747 个样本
- 过滤掉无效域名、良性域名、反向授权区域的域名

- 使用 AVClass 对恶意样本进行分类
- PUP/恶意软件家族分类
- e2LD 分类

恶意软件和 PUP 演变

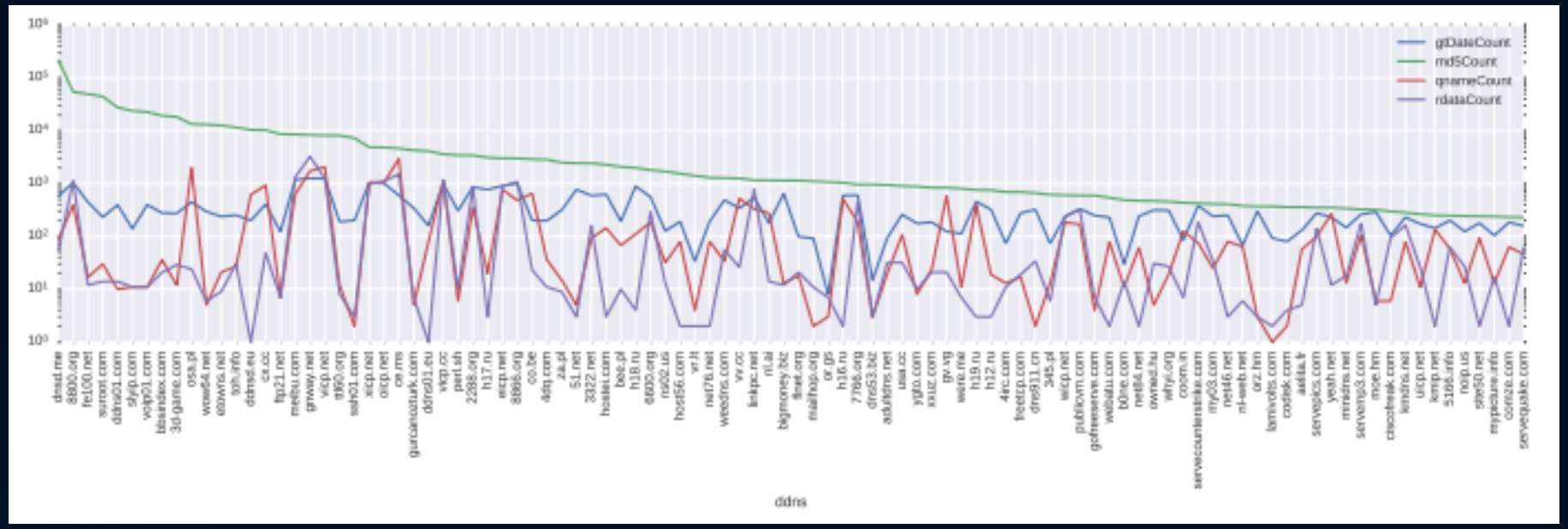


- 恶意软件家族种类多，PUP 单个家族样本数量大
- PUP 的高度多态性可能是为了逃避反病毒引擎的检测

动态恶意软件分析

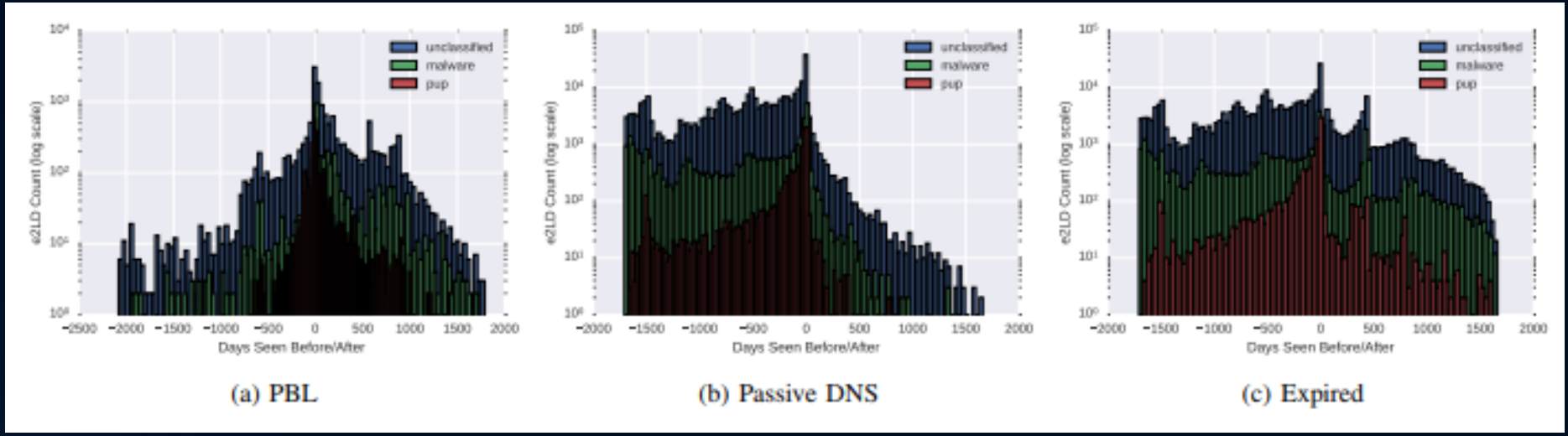
Rank	Family	Samples	Type	e2LDs	FSeen
1	vobfus	2.8 M	Malware	741	11/09
2	multiplug	2.4 M	PUP	808	01/13
3	loadmoney	1.6 M	PUP	2,958	12/12
4	virut	1.4 M	Malware	40,705	03/08
5	softpulse	1.3 M	PUP	3,793	06/14
6	hotbar	1.1 M	PUP	306	08/10
7	installerex	847 K	PUP	155	12/11
8	firseria	795 K	PUP	3,138	07/12
9	outbrowse	771 K	PUP	52	04/13
10	installcore	661 K	PUP	1,118	09/11
Top 10		49%	-	15%	-

Rank	Family	e2LDs	Type	Samples	FSeen
1	virut	40,705	Malware	1.4 M	03/08
2	rodecap	17,382	Malware	11.8 K	05/09
3	zbot	12,959	Malware	163 K	01/08
4	tedroo	6,272	Malware	5 K	11/08
5	sality	4,964	Malware	463 K	12/08
6	upatre	4,658	Malware	503 K	09/13
7	fareit	4,217	Malware	61 K	10/11
8	softpulse	3,793	PUP	1.3 M	06/14
9	ircbot	3,635	Malware	28.5 K	05/06
10	firseria	3,138	PUP	795 K	07/12
Top 10		31%	-	17%	-

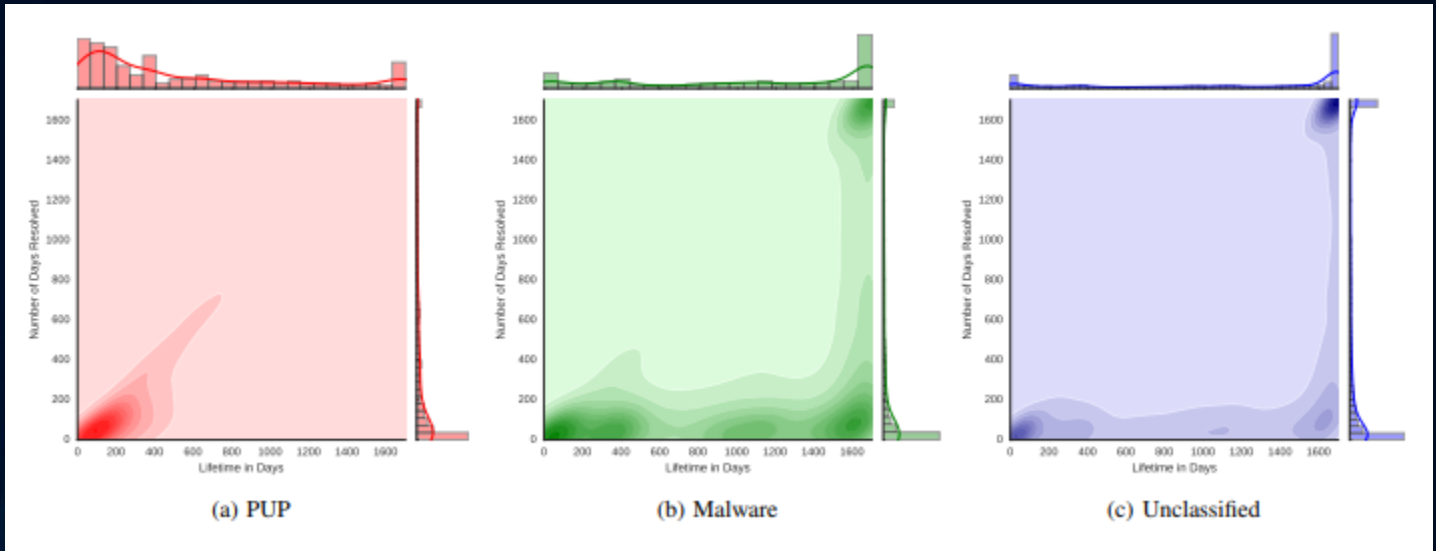


- 域名多态性，大多数MD5解析少于10个独特的e2LD，大多数这些e2LD在五年之间只出现过一次，意味着它们只被单个恶意软件样本查询
- 动态DNS，通过使用动态DNS提供商提供的域名，除非阻断其他合法用户，否则不能在块（zone）这一层阻断恶意使用（最受欢迎的是dnsd.me）
- 内容分发网络使得恶意内容有效隐藏在普通站点中

Passive DNS 与黑名单



- 在动态分析数据集中出现前就进入黑名单的只有30%，20%被延迟超过500天才进入黑名单，依靠Passive DNS信誉系统可以比公开黑名单更快地识别新的威胁
- 动态分析样本之前，PUP相关域名平均出现192天，流行的恶意软件平均延迟623天



(a) PUP (b) Malware (c) Unclassified

- PUP 和恶意软件具有显著不同的 DNS 解析特征：PUP 往往有大量域名生命周期很短，而且很少解析，而恶意软件的生命周期往往很长，且经常被解析

- 我们的数据集占66个可识别DGA中的42个，只有1.8%的DGA域名可以成功解析

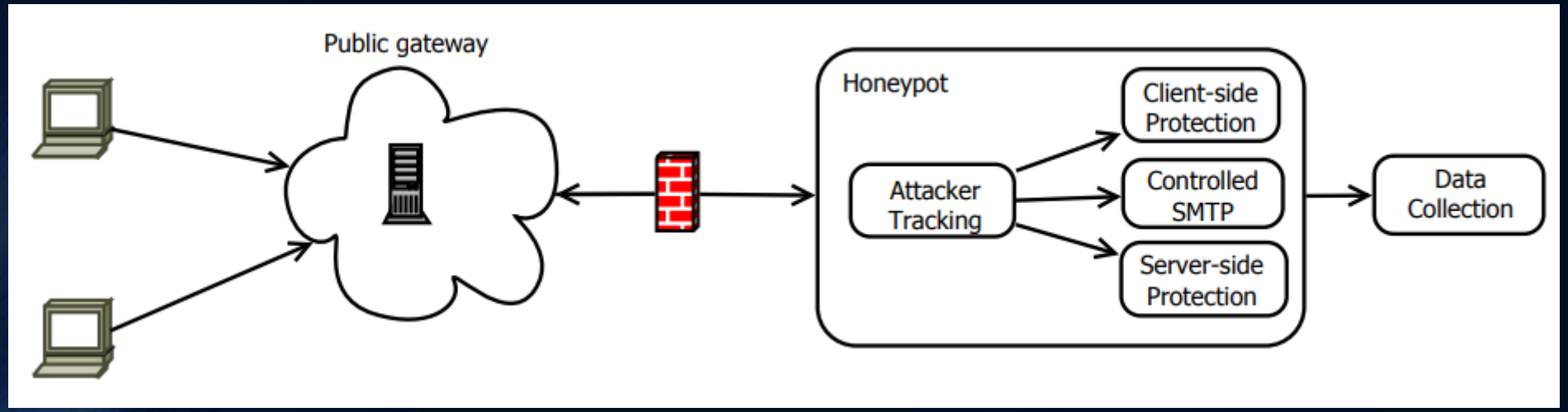
#	Family	Before Filtering	After Filtering
1	virut	2,477,628	40,452
2	pykspa	189,644	180
3	necurs	110,092	1
4	suppobox	72,476	4,677
5	tinba	52,463	682
6	gameover	24,325	7,083
7	emotet	23,500	96
8	pushdo	13,170	17
9	ranbyus	12,922	7
10	nymaim	12,490	148
11	simda	12,348	590
12	murofet	9,295	20
13	qakbot	4,130	119
14	ramnit	3,560	418
15	cryptolocker	2,912	89
16	conficker	1,710	465
17	sisron	1,394	1
18	oderoor	622	3
19	matsnu	525	130
20	dircrypt	510	53
21	tempedreve	204	20
22	banjori	200	1
23	feodo	192	13
24	urlzone	77	18
25	tsifiri	59	58
26	torpig	53	2
27	ramdo	49	27
28	gspy	49	0
29	bamital	48	2
30	bedep	44	5
31	hesperbot	37	2
32	fobber	31	2
33	gozi	24	8
34	bobax	23	0
35	proslikefan	12	1
36	darkshell	10	3
37	redyms	2	2
38	xxhex	1	1
All		3,026,831	55,396

PhishEye: Live Monitoring of Sandboxed Phishing Kits

沙盒钓鱼工具套件的实时监测

两个困境

- 大多数网络钓鱼工具只能在被反钓鱼服务检测到后才被监控
- 没能观测到真正受害者和钓鱼工具包进行交互的方式



五大目标

- Paypal、Apple、Google、Facebook
- French online tax payment system

数据收集

- 收集 643 个独特的网络钓鱼套件
- 初始数据集之外的 474 个
- 127 个网络钓鱼套件连接了 2468 个受害者

攻击者模式

- 29% 基于搜索引擎跳转
- 40% 以上从 facebook 跳转
- 安装后有 70% 的攻击者会浏览查看页面
- 58% 的攻击者会提交虚假凭据验证可行
- 发送钓鱼邮件和网络钓鱼运营维护是解耦的

受害者

- 过往高估了受害者的数量
- 98% 都可以被黑名单最终检出
- 有受害者的会在安装 20 天后被列入黑名单
- 无受害者的会在安装 10 天后被列入黑名单
- 62% 的页面已经出现了 ¾ 的受害者后才被列入黑名单
- 27% 的页面已经出现了 ¼ 的受害者前就被列入黑名单
- 时间分布存在两种类型：偏右分布、双峰分布

TLS in the wild: An Internet-wide analysis of TLS-based protocols for electronic communication

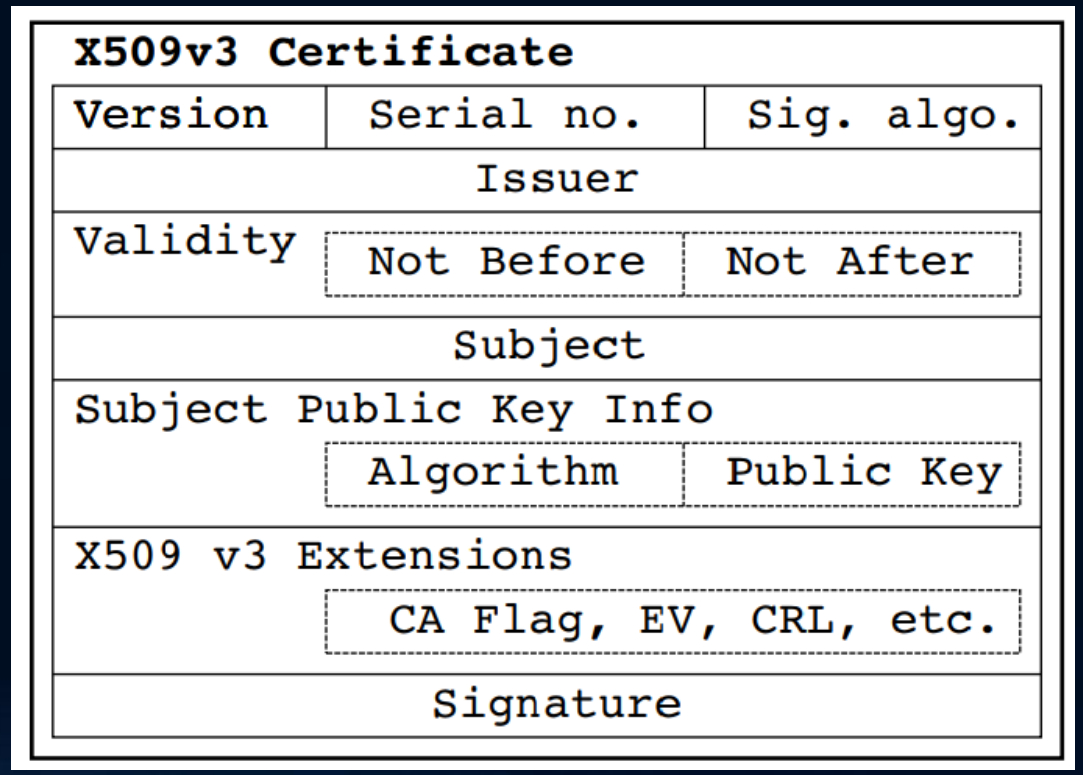
互联网范围基于 TLS 的通信协议分析

整体情况

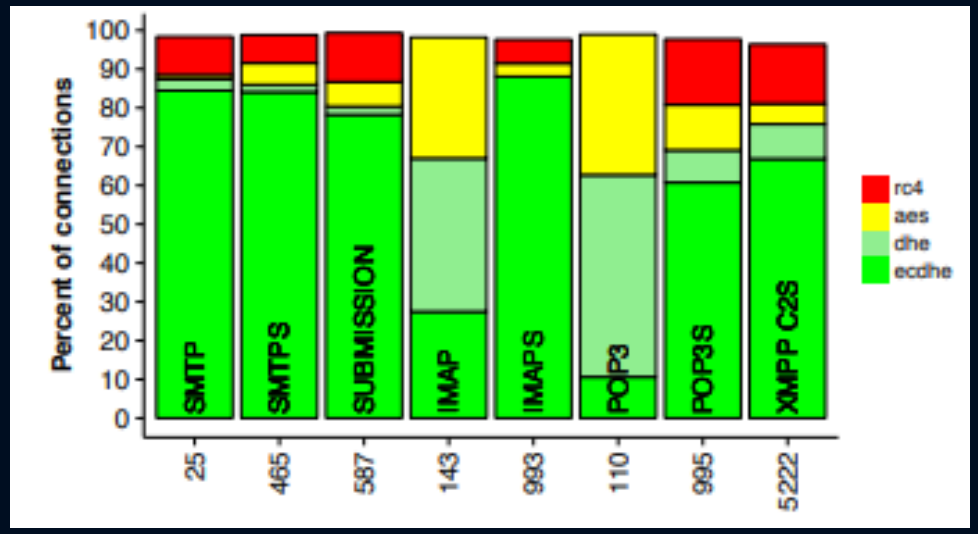
- 用于电子邮件：SMTP / STARTTLS，SMTPS，SUBMISSION，IMAP / STARTTLS，IMAPS，POP3 / STARTTLS 和 POP3S
- 用于聊天：IRC / STARTTLS，IRCS，XMPP / STARTTLS和XMPPS
- 扫描整个 IPv4 地址空间，32 亿路由地址，超过 5000 万活动端口
- 使用 zmap 进行主动扫描，Bro 进行被动监控
- IMAPS 40%，SMTP 40%，XMPPS 27%
- 除大型提供商外的邮件服务大多不安全

X.509 证书

Version	Active probing Negotiated with server	Passive monitoring Observed connections
SSL 3	0.02%	1.74%
TLS 1.0	39.26%	58.79%
TLS 1.1	0.23%	0.1%
TLS 1.2	60.48%	39.37%



普查情况



- 互联网中有许多主机不完成 TCP 握手就回复 SYN 数据包，这也是导致 SSL/TLS 握手失败的原因
- 几乎所有主机都使用 1024 位的，但不建议使用 2048 位以下的
- XMPPS 48% 用于 CDN，12% 用于苹果的推送服务，8% 用于三星的推送服务
- 互联网中证书重用十分普遍
- 99% 的 SUBMISSION 和 90% 的 IMAPS 服务器的密码使用明文传递

Thank you very much !

Information Security Laboratory
North China University of Technology