软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2017,28(9):2370−2387 [doi: 10.13328/j.cnki.jos.005186]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

面向主干网的 DNS 流量监测*

张维维[1,2,3], 龚 俭[1,2,3], 刘尚东[1,2,3], 胡晓艳[1,2,3]

[1](东南大学 计算机科学与工程学院,江苏 南京 210096)

[2](江苏省计算机网络重点实验室,江苏 南京 210096)

[3](计算机网络和信息集成教育部重点实验室(东南大学),江苏 南京 210096)

通讯作者: 张维维, E-mail: wwzhang@njnet.edu.cn

摘 要: 面对 ISP 主干网,为了检测威胁其管理域内用户安全的僵尸网络、钓鱼网站以及垃圾邮件等恶意活动,实时监测流经主干网边界的 DNS 交互报文,并从域名的依赖性和使用位置两个方面刻画 DNS 活动行为模式,而后,基于有监督的多分类器模型,提出面向 ISP 主干网的上层 DNS 活动监测算法 DAOS(binary classifier for DNS activity observation system).其中,依赖性从用户角度观察域名的外在使用情况,而使用位置则关注区域文件中记录的域名内部资源配置.实验结果表明:该算法在不依赖先验知识的前提下,经过两小时的 DNS 活动观测,可以达到 90.5%的检测准确率,以及 2.9%的假阳性和 6.6%的假阴性.若持续观察 1 周,准确率可以上升到 93.9%,假阳性和假阴性也可以下降到 1.3%和 4.8%.

关键词: DNS 监测;域名检测;上层 DNS 流量;DNS 活动分析;多分类器

中图法分类号: TP393

中文引用格式: 张维维,龚俭,刘尚东,胡晓艳.面向主干网的 DNS 流量监测.软件学报,2017,28(9):2370−2387. http://www.jos.org.cn/1000-9825/5186.htm

英文引用格式: Zhang WW, Gong J, Liu SD, Hu XY. DNS surveillance on backbone. Ruan Jian Xue Bao/Journal of Software, 2017,28(9):2370−2387 (in Chinese). http://www.jos.org.cn/1000-9825/5186.htm

DNS Surveillance on Backbone

ZHANG Wei-Wei[1,2,3], GONG Jian[1,2,3], LIU Shang-Dong[1,2,3], HU Xiao-Yan[1,2,3]

[1](School of Computer Science and Engineering, Southeast University, Nanjing 210096, China)

[2](Jiangsu Provincial Key Laboratory of Computer Network Technology, Nanjing 210096, China)

[3](Key Laboratory of Computer Network and Information Integration of Ministry of Education (Southeast University), Nanjing 210096, China)

Abstract: Focusing on ISP backbone, this paper presents a method to detect malicious activities such as botnets, phishing and spam that threaten user security in the domain by monitoring DNS interaction messages through the network boundary in real time. The method depicts DNS behavior patterns based on dependency and position attribute. Then, the paper proposes a supervised classifier based DNS activity detecting algorithm DAOS (binary classifier for DNS activity observation system). Dependency attribute is used to describe external usage of the domain name from perspective of DNS customer, while position attribute is used to describe resource allocation of records in the zone file. Experimental results show that the algorithm, with a DNS data source in 2 hours, can achieve 90.5% of accuracy,

* 基金项目: 国家科技支撑计划(2008BAH37B04); 国家基础研究发展计划(973)(2009CB320505); 国家自然科学基金(60973123)
Foundation item: National Key Technology Research and Development Program of the Ministry of Science and Technology of China (2008BAH37B04); National Program on Key Basic Research Program of China (973) (2009CB320505); National Natural Science Foundation of China (60973123)

收稿时间: 2016-07-11; 修改时间: 2016-09-04, 2016-11-10; 采用时间: 2017-01-06; jos 在线出版时间: 2017-02-20
CNKI 网络优先出版: 2017-02-20 14:05:57, http://www.cnki.net/kcms/detail/11.2560.TP.20170220.1405.019.html

2.9% of false positive, and 6.6% of false negative without prior knowledge. If the observation is kept for a week, accuracy rises up to 93.9%, false positive and false negative can descend to 1.3% and 4.8%.

Key words: DNS surveillance; domain name detection; upper DNS traffic; DNS activity analysis; multi-classifier

随着 Internet 的日益普及,网络的重要性及其对社会的影响越来越大,网络安全问题也随之突显出来.Internet 设计之初考虑最多的是网络互联,几乎没有考虑到安全问题,导致 Internet 具有许多固有的安全问题;其次,入网计算机数量的增加意味着更多的可攻击对象,尤其是防御力差且性能适中的个人主机和移动设备,降低了攻击成本,提高了攻击收益;同时,网络应用数量的快速增长增加了可利用的软件漏洞,尤其是电子商务的出现,使得传统以恶作剧、破坏为主的黑客更趋于经济利益的获取;此外,黑客间的分工协作和攻击手段的工具化,也使得越来越多的业余团体和个人可以参与到产业链中牟取利益[1].作为当前 Internet 最严重的安全威胁之一,Botnet 提供了一个高度可控的攻击平台,方便攻击者发起各种类型的网络攻击.Incapsula在 2015 年的安全报告中指出,Bot 流量近几年来一直保持在 30%左右[2].APWG(国际反网络钓鱼工作组)在 2016 年 3 月一共检测到 28.9 万个钓鱼网站,相比 2015 年 10 月增加 250%.其中,中国以 4.16%的感染用户数排名第二,且以 51.35%的恶意软件感染率排名第一[3].综上可知:Botnet 和钓鱼网站已经成为 Internet 的主要安全威胁,且中国已经成为安全问题最严重的国家之一.网络安全问题空前严峻,亟待解决.

DNS 作为互联网的重要基础设施,承载着域名与 IP 地址间相互映射的重任,网络中各种应用活动都与之密切相关,如电子邮件、网站服务、及时通信、微博等.与此同时,域名解析服务也成为各类互联网安全威胁的重要工具,如僵尸网络在其扩散与通信中使用 DNS 技术定位 C&C(命令控制服务器),网络钓鱼和恶意代码下载等通过频繁变更域名对应的 IP 地址或 NS 记录隐匿背后真实的服务器.DNS 数据可以作为已有检测系统的补充,用于恶意网络活动的检测[4].与全报文网络流量相比,DNS 数据量相对较小且不加密,可以有效地缓解性能压力,实现大规模网络环境下的实时安全监测.另外,各种网络应用都需要事先请求域名解析获取通信 IP 地址,DNS 数据可以进行事前检测,即,在攻击活动开始之前完成检测.

近年来,通过分析 DNS 活动行为以检测恶意服务的实时流量监测技术得到了广泛地研究.该类算法需要对网络中的 DNS 交互报文进行实时或准实时的 DPI 检测,通过挖掘恶意服务有别于合法服务的 DNS 活动特征,以发现恶意活动及其域名.算法的关键在于检测所依据的活动特征,而可选用的特征测度又与流量监测点所处 DNS 层次密切相关.如图 1 所示,根据监测点所处 DNS 层次的不同,DNS 流量大致可以分成两类.

(1) 下层 DNS 流量,是真实终端用户和本地缓存名字服务器 RDNS 间交互的 DNS 请求和响应报文,可以直接从用户网边界的本地缓存名字服务器 RDNS 获取;

(2) 上层 DNS 流量,是从顶级域名服务器 TLD(或者权威名字服务器 NS)采集的 DNS 交互报文,或者从本地缓存名字服务器 RDNS 上方采集的 DNS 交互报文:前者囊括授权域所管辖的所有域名的全球 DNS 解析请求和响应报文,但是该数据的获取需要拥有这些域名服务器的管理员权限;后者是用户网的本地缓存名字服务器 RDNS 和权威名字服务器 NS(或者顶级域名服务器 TLD)之间交互的 DNS 请求和响应报文.

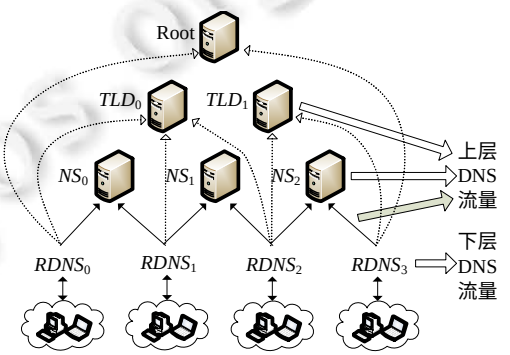


Fig.1 Classification of DNS traffic based on the hierarchy of the DNS in its monitoring points

图 1 DNS 流量按其监测点所处 DNS 层次分类

目前,大量研究工作主要围绕本地缓存名字服务器处获取的下层 DNS 流量开展,从用户角度观测域名的 DNS 活动行为特征.相关工作有:Chatzis 等人依据“邮件蠕虫感染主机的 DNS MX 流量在传播地址和流量特征方面具有高度相似性”提出一系列邮件蠕虫检测方法[5−8];Choi 等人通过观测域名查询请求者的群体活动特性

(即,大量僵尸主机在很短的时间间隔内集中查询某个域名),在大规模网络环境下实现对 Botnet(包括变种和未知类型的 Botnet)及其迁移活动的实时检测[9,10];Jiang 等人利用失效的 DNS 查询请求构造域名和查询主机间的二部图,通过图形分解提取图形结构特点来检测各种网络异常活动[11];Yadav 等人同时关注查询成功的域名及其附近查询失效的域名,通过测量二者间的时间相关性以及域名集信息熵的相似性,提出一个快速的 Botnet 检测算法[12].近年来,攻击者为逃避检测,开始使用多域名、异步访问、僵尸分组等策略加大检测难度.为此,Lee 提出域名查询序列相关性的概念,并基于此构建域名关系图.而后,通过对图形进行边切割划分域名组,以组为单位查询黑名单进行检测.该算法可以有效地检测使用非集中式通信模型的 Botnet[13,14].

本文为了保障 ISP 主干网运行安全,需要实时检测流经主干网边界的上层 DNS 流量.上层 DNS 流量的客户端是 ISP 各个用户网的本地缓存名字服务器 RDNS,而非真实终端用户.为了提高域名查询效率和缓解网络带宽开销,本地缓存名字服务器在一次成功域名解析后的 *TTL* 时间段内,会继续使用其缓存记录向下层用户网提供域名解析服务,而不是重新查询权威名字服务器.DNS 缓存机制会屏蔽下层终端用户的 IP 地址,聚合它们的解析请求,使得上述从终端用户的角度测量域名访问模式相似性的下层 DNS 流量检测算法失效.另一方面,现有针对上层 DNS 流量的检测算法相对较少,代表工作有 Kopis[15]和 Tomas[16].这些算法的基本思路都是基于顶级域名服务器处采集的上层 DNS 流量,从本地缓存名字服务器 RDNS 的视角观察 DNS 访问模式,借用下层 DNS 流量分析中相同的检测模型,通过测量缓存名字服务器的 IP 地址集相似性、空间分布以及查询时间相关性等特征进行检测.但是从顶级域名服务器处可以观察到域名全球范围的 DNS 解析请求和响应报文,即,进行 DNS 查询的缓存名字服务器数量庞大且地域分布广泛.而本文从 ISP 主干网边界采集的上层 DNS 流量,其所辖的本地缓存名字服务器数量少且相对集中,从而降低了上述基于 IP 地址集相似性和空间分布特征的上层 DNS 流量分析手段的准确率.

除 DNS 访问模式外,域名资源记录信息也可用于检测上层 DNS 流量中的恶意活动.僵尸网络和钓鱼网站等恶意服务越来越多地使用 FFSN 来隐匿背后真实的服务器站点(mothership),具有显著的 IP-Flux 和 NS-Flux 活动特征,分别对应域名资源记录“*A*”和“*NS*”.相关工作有:Holz 从垃圾邮件语料库中提取域名,使用 dig 工具主动测量 *A* 记录数目、*NS* 记录数目、*A* 记录所属自治系统数目,而后构建线性分类函数检测 FFSN[17];Caglayan 同时进行主动和被动 DNS 流量采集,选取 *TTL* 值、*A* 记录数目及其空间分布这 3 方面测度,使用贝叶斯置信网络构建分类器,实时检测 FFSN[18].与此同时,ICP 出于负载均衡的考虑,会在全球范围内布署节点服务器构建 CDN 网络,并使用 DNS 轮询机制分发用户请求,同样表现出 IP-Flux 特征.因此,虽然 IP-Flux 和 NS-Flux 作为检测恶意 FFSN 最显著的 DNS 活动特征,却很难区分开合法的 CDN 网络.

另外,为了提高上层 DNS 流量检测算法的准确率,可以综合考虑多方面的 DNS 活动特征.如 Antonakakis 结合域名字面特征、资源记录特征以及证据特征,提出一个动态信誉计算系统 Notos,实时评估给定域名的信誉度[19].Bilge 同时统计域名查询请求的时间分布、域名映射 IP 地址的空间分布、*TTL* 时间长短以及域名字面特征,并应用有监督的决策树算法 J48,提出一个通用域名检测系统 Exposure[20,21].

本文研究适用于 ISP 主干网的上层 DNS 活动监测技术,通过融合主干网边界采集的上层 DNS 流量和流测量数据,综合考虑时间和空间两个维度,从域名的依赖性和使用位置两个方面观察域名的 DNS 活动特征.其中:域名依赖性通过测量用户规模和活跃度,衡量域名对用户的重要程度;域名使用位置通过统计域名资源记录“*A*”和“*NS*”的 IP-Flux 和 NS-Flux 特征,评估用于承载服务的网络基础架构的可靠性.而后,应用有监督的机器学习模型,提出一个适用于上层 DNS 流量的实时检测算法,能够同时兼顾恶意域名检测的及时性和准确率.

1 上层 DNS 活动规律

为了保障 ISP 主干网运行安全,本文基于主干网边界采集的 DNS 交互报文,研究适用于上层 DNS 流量的恶意域名检测算法.其中,算法的关键在于检测所依据的特征测度.为了寻找适用于上层 DNS 流量检测的域名活动特征,本文连续 3 个月(2015 年 3 月 1 日~5 月 31 日)采集流经江苏省教育和科研计算机网(JSERNET)边界的 DNS 交互报文.选取 Alex[22]网站连续 3 次排名前 1 万的域名以及僵尸网络[23,24]、钓鱼网站[25]、垃圾邮件[26]和

恶意软件[23,24]黑名单中出现过的恶意域名,从域名的依赖性和使用位置两个方面观察用户使用域名的外在活动规律以及域名的内部资源部署.为了后续形式化描述的方便,首先给出本文的符号定义.

1.1 符号定义

(1) 域名集合 $D=\{d_1,d_2,d_3,\ldots,d_{|D|}\}$

由于同一个二级域名通常对应 ICP 的一组相关服务,因此,论文以二级域名作为 DNS 流量监测的基本对象,整个待观测域名集合包含$|D|$个二级域名对象 $d_l(1\le l\le|D|)$.

$\forall d_l\in D,1\le l\le|D|,CD_l=\{d_{lm}|1\le m\le|CD_l|\}$表示域名 d_l 所辖所有子域名集合.

(2) DNS 请求集合 $Q=\{q_1,q_2,q_3,\ldots,q_{|Q|}\}$

$\forall q_i\in Q,1\le i\le|Q|,q_i=\langle t,c,s,d,type\rangle$表示本地缓存名字服务器 c 在 t 时刻向上层权威名字服务器(或顶级域名服务器)s 发送域名 d 的 $type$ 类查询请求.

(3) DNS 响应集合 $R_1=\{r_1,r_2,r_3,...,r_{|R_1|}\}$

$\forall r_i\in R_1,1\le i\le|R_1|,r_i=\langle t,c,s,d,type,rcode,aa_flag\rangle$表示上层权威名字服务器(或顶级域名服务器)$s$ 在 t 时刻回复本地缓存名字服务器 c 对于域名 d 的 $type$ 类查询的解析结果,其中,$rcode$=0 表示解析成功,否则为解析失败;aa_flag 为授权回答标志位,aa_flag=1 表示 s 为权威名字服务器.

(4) A 类查询的域名解析结果集合 $R_2=\{r_1,r_2,r_3,...,r_{|R_2|}\}$

$\forall r_i\in R_2,1\le i\le|R_2|,r_i=\langle t,d,a\rangle$表示 t 时刻域名 d 的 A 类查询结果为 a.

(5) NS 类查询的域名解析结果集合 $R_3=\{r_1,r_2,r_3,...,r_{|R_3|}\}$

$\forall r_i\in R_3,1\le i\le|R_3|,r_i=\langle t,d,ns\rangle$表示 t 时刻域名 d 的 NS 类查询结果为 ns.

(6) 流摘要记录集合 $F=\{f_1,f_2,f_3,\ldots,f_{|F|}\}$

$\forall f_k\in F,1\le k\le|F|,f_k=\langle ftime,ltime,sip,sport,cip,cport,proto,pkts,octets\rangle$表示从 $ftime$ 到 $ltime$ 时间段,主机 sip 使用 $sport$ 端口与主机 cip 的 $cport$ 端口进行单方向通信($sip\rightarrow cip$),协议号为 $proto$,报文数目为 $pkts$,字节数为 $octets$.

(7) 时间参数 W(当前时间窗口),W'(历史时间窗口),W_{pre} 和 W_{post} 分别表示前后两个时间窗口.

1.2 域名依赖性

域名访问是指用户发出的关于该域名解析请求的行为,为了检测域名访问异常,本文提出域名依赖性的概念.依赖性是从域名解析请求的用户角度,通过测量用户规模和活跃度,衡量域名对用户的重要程度.但是 DNS 缓存机制的存在,会屏蔽终端用户的 DNS 查询请求.值得庆幸的是,网络流测量数据(如 NetFlow 流数据)中详细地记录了主机与主机之间的通信会话摘要,可以用于弥补上层 DNS 流量中终端用户不可见的不足.如图 2 所示,终端用户在访问目标服务器(如百度搜索引擎)之前,需要根据其域名(baidu.com)查询该服务器的 IP 地址(180.149.132.47),而后才能与该 IP 地址建立通信连接,最终获得所需服务.

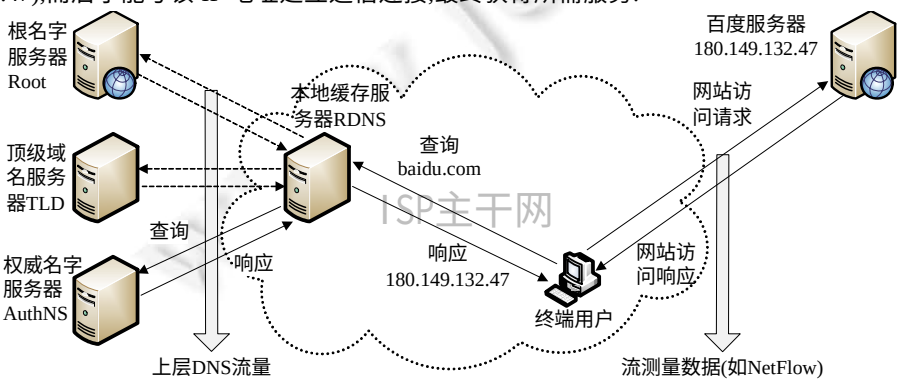


Fig.2 The entire interaction process of users accessing Baiduweb site

图 2 终端用户访问百度网站的整个交互过程

在上述终端用户的整个访问过程中,上层DNS流量可以提供域名和所绑定服务器IP间的映射关系(域名⇔服务器IP),而流测量数据又可以给出终端用户与服务器之间的通信关系(终端用户⇔服务器IP).因此,结合主干网边界采集的上层DNS流量和流测量数据,可以最终得到域名与终端用户间的关联关系(域名⇔服务器IP⇔终端用户),进而统计域名依赖性的相关测度.

1.2.1 用户规模

(1) 用户IP地址数目 *ipNum*:

$$resolvedIpSet(d_l,W)=\{a(r_i)|\forall r_i\in R_2,d(r_i)=d_l \text{ 且 } t(r_i)\in W\} \tag{1}$$

$$\begin{aligned} ipSet(d_l,W)=&\{cip(f_k)\mid \forall f_k\in F, sip(f_k)\in resolvedIpSet(d_l,W)\text{且}ltime(f_k)\in W\}\cup\\ &\{sip(f_k)\mid \forall f_k\in F, cip(f_k)\in resolvedIpSet(d_l,W)\text{且}ltime(f_k)\in W\} \end{aligned} \tag{2}$$

$$ipNum(d_l,W)=|ipSet(d_l,W)| \tag{3}$$

(2) 用户逻辑归属分布:平均每个单位用户IP数 *ipNumPerUnit*;平均每个单位的用户IP数目占该单位活跃IP总数的比重 *ipRatioPerUnit*:

$$unitSet(d_l,W)=unit(ipSet(d_l,W))=\{unit(ip)|\forall ip\in ipSet(d_l,W)\} \tag{4}$$

$$unitNum(d_l,W)=|unitSet(d_l,W)| \tag{5}$$

$$ipNumPerUnit(d_l,W)=\frac{ipNum(d_l,W)}{unitNum(d_l,W)} \tag{6}$$

$$ipRatioPerUnit(d_l,W)=\frac{\sum_{u\in unitSet(d_l,W)}\frac{|\{ip\mid \forall ip\in ipSet(d_l,W)\text{且}unit(ip)=u\}|}{ipSumOfUnit(u)}}{unitNum(d_l,W)} \tag{7}$$

(3) 用户地理位置分布:平均每个城市的用户IP数 *ipNumPerCity*:

$$citySet(d_l,W)=city(ipSet(d_l,W))=\{city(ip)|\forall ip\in ipSet(d_l,W)\} \tag{8}$$

$$cityNum(d_l,W)=|citySet(d_l,W)| \tag{9}$$

$$ipNumPerCity(d_l,W)=\frac{ipNum(d_l,W)}{cityNum(d_l,W)} \tag{10}$$

用户规模通过统计访问域名的用户IP地址数目、逻辑归属分布和地理位置分布,从空间角度测量域名的依赖性.作为评估域名依赖性的基础测度,用户IP地址数目的统计需要结合域名 *A* 类查询的解析结果和流测量数据.在 *W* 时间窗口内,首先根据域名检索其 *A* 类查询解析结果,得到所有解析IP地址集合(公式(1)),而后查询流测量数据,找到这些解析IP地址所有通信会话的对端IP集合(公式(2)),用户IP地址数即为该集合的大小(公式(3)).如图3所示,僵尸网络域名拥有最大的用户群,其中,75.7%的域名拥有10个以上用户;也有25.1%的合法域名,其用户数超过10;而只有3.7%的钓鱼网站、9.6%的垃圾邮件以及13.4%的其他恶意软件域名拥有相同数目的访问用户.观察结果表明:用户IP地址数可以区分开合法域名和恶意域名,尤其是僵尸网络域名.

用户IP地址数目仅能反映总量的多少,为了细粒度地刻画用户的空间分布规律,本文从逻辑归属和地理位置两个空间维度,分别观察用户空间范围和各个子空间的用户密度.传统DNS流量分析中提出组织机构数目(如BGP前缀数、自治系统AS数、ISP数和单位数)和国家地区数目(如国家数、城市数)作为测量空间范围的二级测度.但是对于ISP主干网而言,由于用户局限于其所在的ISP、AS和国家,因此只需要选取其中的单位数目(公式(5))和城市数目(公式(9)).实验观察发现:各类域名按单位数目或者城市数目的累积分布曲线与图3相似,僵尸网络域名的用户群空间分布最广,与其他类别域名的曲线区别显著,但是合法域名与垃圾邮件域名的两条累积分布曲线距离相差只有10%左右,即,从空间范围很难区分开这两类域名.

为此,本文提出二级测度平均每个单位(或者城市)的用户数(公式(6)和公式(10)),用以测量子空间密度.观察发现:平均每个单位的访问用户数超过1.5的僵尸网络域名有83.1%;而只有42.9%的合法域名、11.5%的钓鱼网站域名、14.1%的垃圾邮件域名和27.8%的其他恶意软件域名,其平均每个单位的访问用户数超过1.5.若不考虑其他恶意软件域名,则合法域名和垃圾邮件域名(或钓鱼网站域名)的两条累积分布曲线的距离相差接近

30%,具有明显的差异.此外,考虑到各个子网拥有的 IP 地址数目差别较大,大的用户子网可能包含几个 *B* 类地址,而小的用户子网只有 1~2 个 *C* 类地址,若对所有单位计算平均用户数则不公平.因此,进一步提出二级测度平均每个单位的用户 IP 数占该单位活跃 IP 总数的比重(公式(7))来测量子空间密度.实际观察结果表明(如图 4 所示):考虑各个用户网的实际活跃 IP 数后,当平均每个单位的用户比重超过 10^{-4} 时,僵尸网络域名数占总数的累积比重为 84.6%,而合法域名数占总数的累积比重仅为 29.9%,相差 54.7%,即能够更加准确地识别僵尸网络域名.

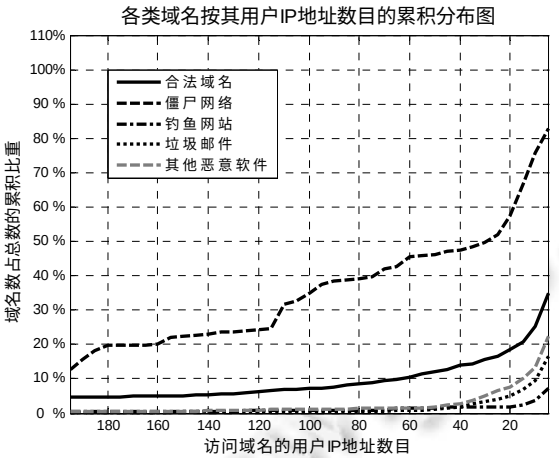


Fig.3 Cumulative distribution of domain names according to their user numbers

图3 域名按其用户 IP 地址数目的累积分布图

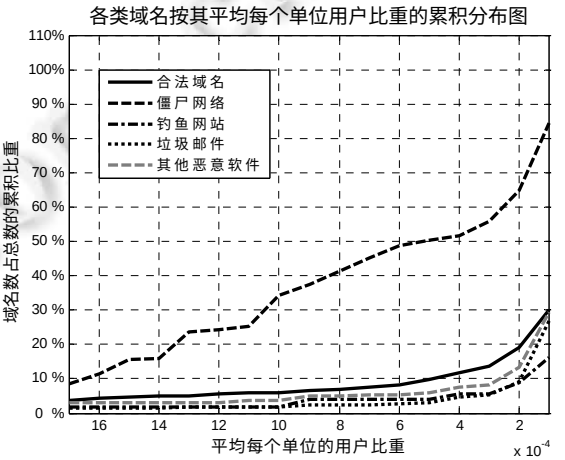


Fig.4 Cumulative distributionof domain names according to the average users count ratio of each unit to the whole network

图4 域名按其平均每个单位用户比重的累积分布图

1.2.2 活跃度

(1) 域名生命周期 *liveSpan*:

$$liveSpan(d_l)=\lceil (ltime(d_l)-ftime(d_l))/3600 \rceil \tag{11}$$

(2) 域名活跃频率 *clientAppearFreq*:

$$clientAppearFreq(d_l)=\frac{|\{h \mid \forall 0 \le h < liveSpan(d_l), ipNum(d_l,h)>0\}|}{liveSpan(d_l)} \tag{12}$$

(3) 周期性 *clientLineDistance*:

$$clientLineDistance(d_l)=\sum_{0\le h<24}\left(\frac{ipNum(d_l,h+24)-ipNum(d_l,h)}{ipNum(d_l,h)}\right) \tag{13}$$

(4) 用户群稳定性 *clientDiffRatio*:

$$clientDiffRatio(d_l,W_{pre},W_{post})=1-\frac{|ipSet(d_l,W_{pre})\cap ipSet(d_l,W_{post})|}{|ipSet(d_l,W_{pre})\cup ipSet(d_l,W_{post})|} \tag{14}$$

(5) 用户访问突发性 *abruptRatio*:

$$abruptRatio(d_l,W)=1-\frac{Avg(ipNumSereis(d_l,\ ftime(d_l),ltime(d_l)))}{Max(ipNumSereis(d_l,\ ftime(d_l),ltime(d_l)))} \tag{15}$$

(6) 解析失败率 *failureRatio*:

$$failureTimes(d_l,W)=|\{r_i|\forall r_i\in R_1, d(r_i)=d_l \text{ 且 } t(r_i)\in W \text{ 且 } rcode(r_i)\neq 0\}| \tag{16}$$

$$failureRatio(d_l,W)=\frac{failureTimes(d_l,W)}{|\{r_i \mid \forall r_i\in R_1, d(r_i)=d_l 且 t(r_i)\in W\}|} \tag{17}$$

活跃度关注域名的存活时间和活动频率,从时间角度测量域名的依赖性.针对长期存活的域名,探查域名活

动的周期性和活动频率;而对于短期存活的域名,则通过观察用户群稳定性,用户访问突发性以及 Domain- Flux 特征,研究域名活动的社区通信特征.此处需要重点说明的是:传统 DNS 流量分析工作中,对于周期性和突发性特征的统计,大多都是建立在对 DNS 查询次数时间序列的统计上.但是 DNS 缓存机制的存在,使得下层用户的域名解析请求,只有在缓存中没有该资源记录或者 *TTL* 超时的情况下,才会出现在上层 DNS 流量中.即:在上层 DNS 流量分析中,对 DNS 查询次数的统计将失去原先意义.为此,本文借助流测量数据中提供的主机通信关系,统计域名各个时间片的用户 IP 数.与 DNS 查询次数相比,用户数可以更直观地体现域名的使用量.而后,在用户 IP 数时间序列基础上,进一步分析域名活动的周期性和突发性.

通常情况下,ICP 为了保障用户能够持续获取服务,一般不会变更域名.而恶意服务为了躲避黑名单过滤,则会频繁地变更域名,因此,合法域名的生命周期远远长于恶意域名.如图 5 所示:87.0%的合法域名生命周期长度超过 75 天,只有 5%左右的僵尸网络和钓鱼网站域名以及 15%左右的垃圾邮件和其他恶意软件域名的生命周期长度超过 75 天.同时,观察发现:生命周期不足 1 天的合法域名几乎没有(<0.4%);而生命周期不足 1 天的恶意域名却有一半左右,其中,僵尸网络域名 69.0%、钓鱼网站域名 59.3%、垃圾邮件域名 48.1%、其他恶意软件域名 45.3%.合法域名和恶意域名的生命周期长度存在显著的差异,即,生命周期可以作为一个有效的检测测度使用.

生命周期的长短只能反映域名存活的时间跨度,在这段时间内,域名可能持续活跃,也可能偶尔活跃.为了观察域名的具体活跃情况,进一步统计其活跃频率,即,整个生命周期中活跃时间所占比重(见公式(12)).实际观察发现:僵尸网络域名由于用户的短时间集中访问,拥有最大的活跃频率(活跃频率超过 97.5%的域名有 93.4%);很多合法域名由于长期间歇访问,活跃频率反而较小(只有 28.8%的域名的活跃频率超过 97.5%);而钓鱼网站、垃圾邮件及其他恶意软件域名的活跃频率则介于两者之间.

对于 ISP 主干网而言,由于其用户群属于同一个 ISP 和地理区域,大多数用户每天具有相似的域名访问行为.如:JSERNET 内,各个高校的大学生拥有相同的作息时间,对于知名合法域名的访问具有较强的周期性(T=1 天);而对于钓鱼网站、垃圾邮件等恶意软件域名,只有受害者才会访问,表现出相对较弱的周期性,但是对于僵尸网络而言,僵尸主机的同步性使得其域名访问活动具有更强的周期性.一个简单的周期性检测方法,可以通过计算前后两天用户 IP 地址数时间序列曲线的欧拉距离.但是不同域名所拥有的用户群规模相差较大,使得大用户群域名曲线的微小变化也会超过小用户群域名曲线的较大变化.因此,本文在欧拉距离的基础上,通过比上曲线均值计算相对距离(见公式(13)).实际观察发现(如图 6 所示):超过 0.39 的合法域名有 67.7%,钓鱼网站、垃圾邮件等恶意软件域名有 85%左右,而僵尸网络域名只有 14.8%.

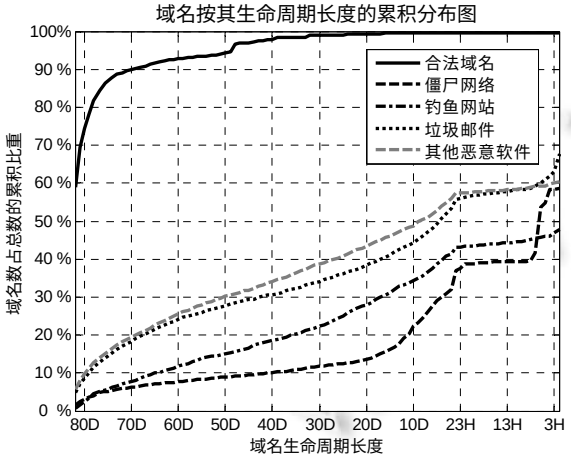


Fig.5 Cumulative distributionof domain names according to their lifespan

图 5 各类域名按其生命周期长度的累积分布图

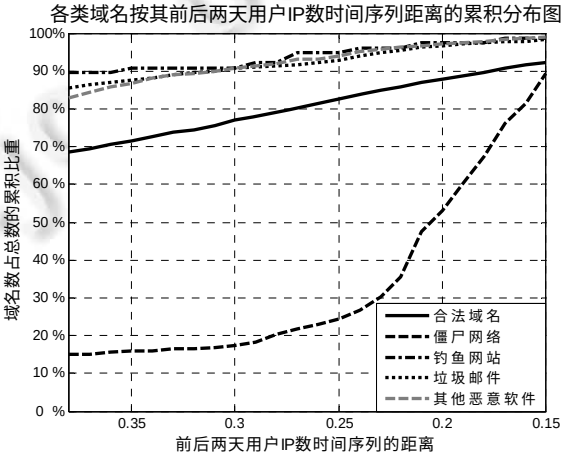


Fig.6 Cumulative distributionof domain names according to the distance between two consecutive days' time series of the user numbers

图 6 各类域名按前后两天用户 IP 数时间序列距离累积分布图

短期活跃的恶意服务,尤其是僵尸网络,通常拥有稳定的用户群、突发的用户访问量以及频繁变更域名所引发的大量 NXDomain 响应报文.为此,本文提出域名活动的社区通信特征:通过统计前后两个时间窗口用户 IP 地址集合的差异度,测量用户群的稳定性(见公式(14));通过测量用户 IP 数时间序列曲线中平均值和最大值的相对距离,评估曲线中存在异常突起的可能性(见公式(15));同时,计量域名的 NXDomain 响应报文数,检测 Domain- Flux 特征(见公式(16)和公式(17)).实际观察结果表明:僵尸网络域名具有最小的用户 IP 地址集差异度(如图 7 所示,其中,差异度超过 97.5%的僵尸网络域名只有 32.1%,而合法域名和其他类型的恶意域名至少有 65.9%),说明僵尸网络拥有更加稳定的用户群;僵尸网络域名的用户 IP 数时间序列曲线存在异常突起的可能性最大(如图 8 所示,可能性不足 0.025 的僵尸网络域名只有 15%,而合法域名和其他类型的恶意域名却超过 52.5%),表明僵尸网络的用户访问具有很强的突发性;僵尸网络域名具有最大的解析失败率(解析失败率不足 2.5%的僵尸网络域名仅有 32.3%,而合法域名和其他类型的恶意域名却超过 93.0%),证明僵尸网络使用的 Domain-Flux 技术会产生大量的 NXDomain 出错响应报文.

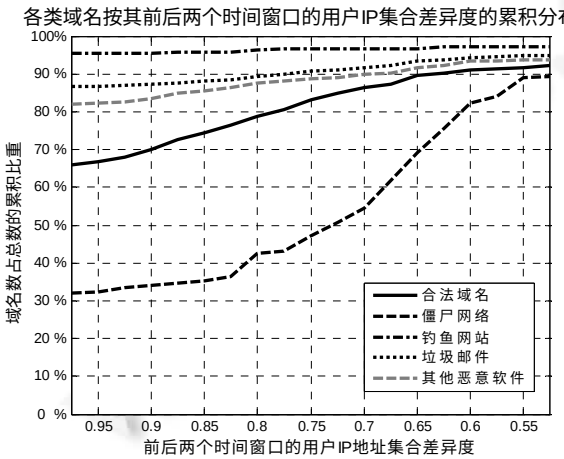


Fig.7 Cumulative distributionof domain names according to the diversity factor of two IP address sets from two consecutive time windows

图 7 各类域名按前后两个时间窗口用户 IP 集差异度累积分布图

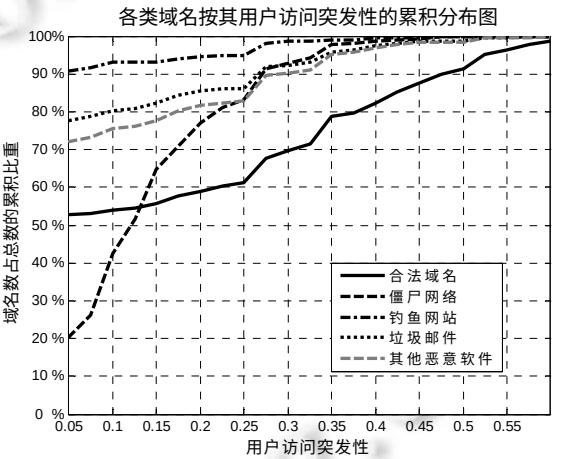


Fig.8 Cumulative distributionof domain names according to the burstiness of their user access

图 8 各类域名按访问突发性的累积分布图

1.3 域名使用位置

资源记录 *A* 和 *NS* 中隐含域名的使用位置信息,前者记录域名的解析 IP 地址,表明哪里可以访问域名绑定的服务;后者记录域名的权威名字服务器,给出哪里可以查询域名的资源记录信息.与 FFSN 相比,CDN[27]虽然也使用轮询机制将域名映射到一组 IP 地址,但是存在根本的区别.

(1) CDN 背后的节点服务器通常是一组高性能专用服务器,拥有的 IP 地址集合相对稳定且长期存活.而 FFSN 背后的僵尸代理主要是受病毒感染的普通用户主机,由于攻击者无法直接控制机器的开关,因此,用作僵尸代理的主机常常短期活跃,使得集合内的 IP 地址不断更替(新僵尸主机的加入或原有僵尸主机的退出);

(2) CDN 为了提高服务质量,减少用户访问的响应时间,尽可能使用户就近获取所需内容.需要在全球范围内均匀部署节点服务器,并指定离用户最近的服务器 IP 地址作为域名查询的解析结果.因此,CDN 域名解析的 IP 地址集空间分布广且均匀,就近原则也使得靠近用户群的服务器 IP 地址被轮询到的概率变大.对于 FFSN 而言,由于同一个组织机构的用户网络存在许多相似的安全漏洞,很可能成片地受到感染,即,攻击者可利用的僵尸代理局部集中分布.此外,攻击者为了避免黑名单过滤,会频繁地将域名映射到新的僵尸代理,随机轮询使得 FFSN 域名各个解析 IP 地址被轮询到的概率相当;

(3) ICP 部署 CDN,通常在其上提供一组相关服务,即:CDN 节点服务器的 IP 地址,常常会反向映射到多个域

名,但是它们具有相同的二级域名;而 FFSN 使用僵尸作为其服务代理,为了提高僵尸主机的使用率,同一个攻击者(或者不同的攻击者)可以租赁同一个僵尸绑定不同的服务,即,FFSN 域名解析的 IP 地址可以反向映射到不同的二级域名.尤其是使用 Domain-Flux 技术的新型僵尸网络,一个 IP 地址可以反向映射到一组随机域名.

为了识别混入 FFSN 的 CDN,降低检测的误报率,本文在传统测量解析 IP 地址数目、权威名字服务器数目、空间分布以及 *TTL* 值检测 IP-Flux 和 NS-Flux 特征的基础上,根据上述分析提出一组新的测度.从解析 IP 地址和权威名字服务器两个方面,通过观察各自对象集合的稳定性、各个对象的存活时间、解析次数分布、反向映射的二级域名数目和 *TTL* 分布等,评估用于承载服务的网络基础架构的可靠性.

1.3.1 IP-Flux 特征分析

(1) 解析 IP 地址集合空间分布特征:各解析 IP 出现次数的标准方差均值比 *ripTimesDistribution*;当前时间窗口所有解析 IP 反向映射的二级域名总数 *ripSdomainTNum*;历史时间窗口平均每个解析 IP 曾经反向映射过的二级域名数 *ripSdomainMnum*:

$$ripTimes(d_l,W,a_m)=|\{r_i|\forall r_i\in R_2,d(r_i)=d_l \text{ 且 } t(r_i)\in W \text{ 且 } a(r_i)=a_m\}| \tag{18}$$

$$ripTimesSeries(d_l,W)=\{ripTimes(d_l,W,a_m)|a_m\in resolvedIpSet(d_l,W)\} \tag{19}$$

$$ripTimesDistribution(d_l,W)=\frac{Var(ripTimesSeries(d_l,W))}{Avg(ripTimesSeries(d_l,W))} \tag{20}$$

$$ripSdomainTNum(d_l,W)=|\{d(r_i)|\forall r_i\in R_2,a(r_i)\in resolvedIpSet(d_l,W)\text{且 } t(r_i)\in W\}| \tag{21}$$

$$ripSdomainMNum(d_l,W')=\frac{\sum_{a_m\in resolvedIpSet(d_l,W)}|\{d(r_i)|\forall r_i\in R_2,a(r_i)=a_m\text{且}t(r_i)\in W'\}|}{|resolvedIpSet(d_l,W)|} \tag{22}$$

(2) 解析 IP 地址集合时间分布特征:前后两个时间窗口解析 IP 地址集差异度 *ripDiffRatio*;解析 IP 地址的平均生命周期长度 *ripLiveMSpan*:

$$ripDiffRatio(d_l,W_{pre},W_{post})=1-\frac{|resolvedIpSet(d_l,W_{pre})\cap resolvedIpSet(d_l,W_{post})|}{|resolvedIpSet(d_l,W_{pre})\cup resolvedIpSet(d_l,W_{post})|} \tag{23}$$

$$ripLiveMSpan(d_l,W)=\frac{\sum_{a_m\in resolvedIpSet(d_l,W)}(\lceil(ltime(a_m)-ftime(a_m))/3600\rceil)}{|resolvedIpSet(d_l,W)|} \tag{24}$$

CDN 的就近原则使得靠近用户群的服务器 IP 地址被轮询到的概率大,FFSN 的随机轮询使得各个解析 IP 地址被轮询到的概率相当.为了观察这一现象,测量解析 IP 地址出现次数的分布特征.作为评估数据集离散程度的重要测度,标准方差的主要思想是计算各个数据偏离平均数的距离.但是对于不同域名,其解析 IP 地址出现次数的均值各不相同,单独统计标准方差,无法直观地刻画不同域名间解析 IP 地址出现次数的分布特征,因此,本文在标准方差计算的基础上进一步比上平均值,提出了标准方差均值比(见公式(20))这一测度,其中,参数 *ripTimesSeries* 表示某个域名对象在时间窗口 W 内的所有解析 IP 地址的出现次数序列(公式(19)).实验观察发现:解析 IP 地址出现次数的标准方差均值比超过 0.6 的 CDN 域名有 53.3%,而 FFSN 域名只有 17.5%.这符合 CDN 的就近访问原则以及 FFSN 的随机轮询机制.

传统的 FFSN 检测算法大多只关注域名→解析 IP 地址的映射关系,而忽略解析 IP 地址→域名的反向映射关系.为了避免域名单点失效,攻击者会使用 Domain-Flux 技术,通过 DGA 算法生成一组随机域名,解析到命令控制服务器.相反,FFSN 命令控制服务器的 IP 地址会反向映射到一组 DGA 域名.而 CDN 节点服务器为了提高自身的利用率,也会在同一台服务器上运行多个服务.即:CDN 节点服务器的 IP 地址集也会反向映射到多个域名,但是它们常常属于同一个或少数几个二级域名.为了观察上述活动现象,本文在统计解析 IP 地址反向映射的域名数[14,15]基础上进一步测量当前时间窗口 W 内域名所有解析 IP 地址反向映射的二级域名总数(见公式(21)).检测算法出于及时性的考虑,W 时间长度通常设置的较短.为了弥补当前数据不足的缺陷,可以使用较长时段的历史相关数据来提高检测准确率.为此,进一步统计历史时间窗口 W' 内域名平均每个解析 IP 地址反向映射的二级域名数(见公式(22)).实际观察发现:所有解析 IP 地址反向映射的二级域名总数超过 39 的 CDN 域名只有

17.3%,而 FFSN 域名却有 42.5%;平均每个解析 IP 地址反向映射的二级域名数超过 10 的 CDN 域名仅有 15.3%,而 FFSN 域名却有 45.4%.与 FFSN 相比,CDN 域名的解析 IP 地址反向映射的二级域名数更少,这符合理论分析的结果.

CDN 部署的节点服务器 IP 地址集合相对稳定且长期存活,而 FFSN 选用的僵尸代理 IP 地址集频繁变更.为了观察这一现象,提出二级测度“前后两个时间窗口解析 IP 地址集合的差异度”(见公式(23))和“解析 IP 地址的平均生命周期长度”(见公式(24)):前者测量解析 IP 地址集合的稳定性,后者关注解析 IP 地址的活跃时间长度.实际统计发现:解析 IP 地址集合差异度超过 97.5%的 CDN 域名有 31.9%,而 FFSN 域名却有 65.8%;解析 IP 地址平均生命周期长度超过 12 天的 CDN 域名有 89.2%,而 FFSN 域名只有 69.2%.充分说明 CDN 域名的解析 IP 地址集合更加稳定,且解析 IP 地址存活时间更久.

1.3.2 NS-Flux 特征分析

(1) 权威名字服务器集合空间分布特征:各名字服务器出现次数的标准方差均值比 *nsTimesDistribution*;当前时间窗口所有名字服务器解析的二级域名总数 *nsSdomainTNum*;历史时间窗口平均每个名字服务解析的二级域名数 *nsSdomainMNum*;

(2) 权威名字服务器集合时间分布特征:前后两个时间窗口名字服务器域名集合的差异度 *nsDiffRatio*;名字服务器 IP 地址集合的差异度 *nsIpDiffRatio*;权威名字服务器的平均生命周期长度 *nsLiveMSpan*.

仿照 IP-Flux 特征测度,设计 NS-Flux 特征测度,具体计算方法类似公式(18)~公式(24),这里不再详细阐述.服务提供商通常会配置一个主 DNS 服务器和多个备用 DNS 服务器,以保证设备在出现故障或者受到恶意攻击时仍然能够继续提供域名解析服务.具体工作时,本地缓存名字服务器在接收到域名请求后会查询首选 DNS 服务器,只有当首选 DNS 服务器无法正常解析域名时才使用备用 DNS 服务器.而 FFSN 使用僵尸主机充当域名解析代理,通过代理的频繁变换,避免受胁迫 DNS 服务器出现单点失效问题.因此,CDN 域名的 *NS* 查询结果比较集中,而恶意 FFSN 域名的 *NS* 返回结果相对均匀.为了描述这一活动现象,提出了域名各权威名字服务器出现次数的标准方差均值比.实际观察发现:各权威名字服务器出现次数的标准方差均值比超过 0.15 的 CDN 域名有 30.1%,而恶意 FFSN 域名只有 14.0%,与理论分析一致.

通常情况下,用户在域名注册商处注册域名,由他们负责域名解析.域名注册商部署 DNS 服务器,对其管理的大量域名提供解析服务,即:CDN 域名映射的权威名字服务器,反向映射到大量二级域名.而恶意 FFSN 会频繁地变更僵尸代理的域名和 IP 地址,使得 FFSN 域名映射的权威名字服务器,反向映射的二级域名数目较少.为了观察这一活动现象,测量当前时间窗口 W 内域名所有权威名字服务器反向映射的二级域名总数,以及历史时间窗口 W'内域名平均每个权威名字服务器反向映射的二级域名数.实际统计可知:所有权威名字服务器反向映射的二级域名总数超过 14 的 CDN 域名有 58.7%,而 FFSN 域名只有 40.1%;平均每个权威名字服务器反向映射的二级域名数超过 3 的 CDN 域名有 67.1%,而 FFSN 域名却只有 49.1%.与 FFSN 相比,CDN 域名映射的权威名字服务器,其反向映射的二级域名数更少,符合理论分析的结果.

最后,从时间维度观察 CND 域名和 FFSN 域名的 NS-Flux 活动特征.为了保证域名解析服务的性能和安全性,域名注册商分布式部署的 DNS 服务器常常是一组高性能的专用服务器,其域名和 IP 地址集合相对稳定且长期存活.而 FFSN 能够选作域名解析代理的僵尸主机主要是受到病毒感染的普通用户主机,因此,FFSN 域名的权威名字服务器通常短期活跃且频繁更替.为此,提出二级测度前后两个时间窗口权威名字服务器域名集合的差异度、前后两个时间窗口权威名字服务器 IP 地址集合的差异度以及权威名字服务器的平均生命周期长度.前两个测度用于测量权威名字服务器集合的稳定性,而后者关注权威名字服务器的活跃时间长度.实验观察发现:前后两个时间窗口权威名字服务器的域名集合差异度超过 62.5%的 CDN 域名只有 4.7%,而 FFSN 域名有 18.7%;前后两个时间窗口权威名字服务器的 IP 地址集合差异度超过 97.5%的 CDN 域名只有 20.0%,而 FFSN 域名却有 89.5%;权威名字服务器平均生命周期长度超过 13 天的 CDN 域名有 95.0%,而恶意 FFSN 域名只有 73.1%.观察结果表明:与恶意 FFSN 域名的权威名字服务器相比,CDN 域名的权威名字服务器的域名和 IP 地址集合更加稳定,且权威名字服务器的存活时间更久.

2 DNS 流量监测模型

面对 ISP 主干网,为了检测威胁其管理域内用户安全的僵尸网络、钓鱼网站以及垃圾邮件等恶意活动,本文提出一种适用于 ISP 主干网的上层 DNS 流量监测模型 DAOS.该算法实时监测流经主干网边界的 DNS 流量,从时间和空间两个维度观察域名依赖性和使用位置两方面的 DNS 活动特征,并分别使用有监督的机器学习算法进行检测.而后,通过加权平均融合两方面检测结果,及时准确地识别恶意域名.

如图 9 所示,整个 DNS 流量检测模型包括 4 个主要功能模块:首先,通过域名聚类算法将待处理的标准域名集和实测域名集各自划分成组,以组为最小单元进行观察和检测;其次,针对数量庞大的测度集进行测度选取,消除冗余测度以及特征不明显的测度,降低计算复杂度和系统开销;而后,对每一组域名,基于 ISP 主干网边界实时采集的 DNS 流量和流测量数据,统计所选测度集中的每一个测度;最后,设计多分类器检测模型,从域名依赖性和域名使用位置两个方面,分别应用有监督的机器学习方法,通过标准域名集的训练学习,检测实测域名集中出现的恶意域名.并在此基础上进一步使用加权平均的数学统计方法,对上述两方面的检测结果进行数据融合,以提高最终检测结果的准确率.

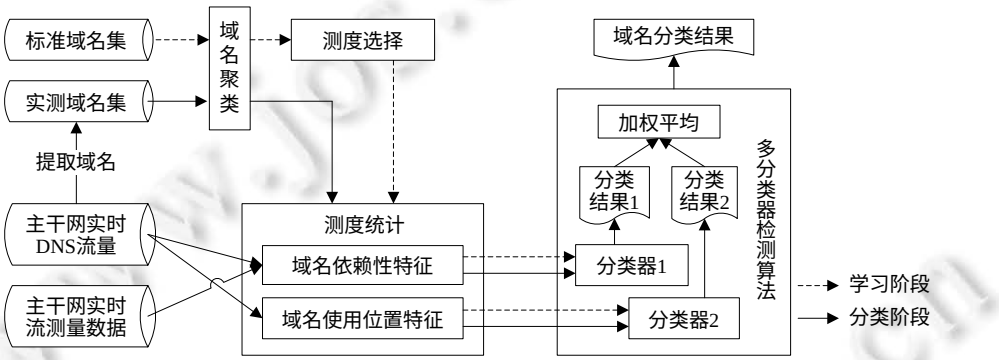


Fig.9 System architecture ofa DNS traffic detection algorithm

图 9 上层 DNS 流量监测模型 DAOS

2.1 域名聚类算法

面对数量庞大的域名集合,若对每一个域名都进行 DNS 活动监测,则系统开销太大.域名聚类的目标在于尽可能地将同一个应用或同一组相近应用划分到一个组,以组为基本单元,监测其 DNS 活动轨迹.常用域名聚类算法主要有 3 种.

(1) 将具有相同二级域名的所有子域名划分成一组,如 XXX.taobao.com.一个单位或者组织机构通常会注册一个二级域名,并在其下命名一组子域名,用以绑定不同的服务或者访问资源.因此,该域名聚类算法能够有效地聚合同一个单位或者个人提供的一组相关服务;

(2) 按解析 IP 地址分组域名,即,将一个解析 IP 地址反向映射的所有域名划分到一个分组,该域名聚类算法能够有效地聚合同一台物理设备上运行的一个或一组服务;

(3) 构建域名集和解析 IP 地址集间的二部图$\langle D,S,E\rangle$(D 表示域名集合,S 表示解析 IP 地址集合,$\forall e=(d,s)$,表示域名 d 的解析 IP 地址为 s),将同一个连通二部图内的所有域名划分成一组.该域名聚类算法可以有效地发现同时使用 IP-Flux 技术和 Domain-Flux 技术的僵尸命令控制器.

为了选择合适的域名聚类算法,本文从正确分组比例 $accrurate(G_i)=\max(g(G_i)/|G_i|,b(G_i)/|G_i|)$和压缩比例 $compression_ratio=1-|G|/|D|$两个方面进行评估.首先,对于一个聚类算法而言,需要尽可能地将同类对象划分到一个分组,不同类对象划分到不同分组.为了计算正确分组比例,分别计算分组中合法域名所占比例以及恶意域名所占比例,而后,选择两者中较大的一个作为最终结果(其中,$g(G_i)$和 $b(G_i)$分别表示分组 G_i 中的合法域名数和恶意域名数).如图 10 所示,3 种域名聚类算法都能将 99%以上的分组进行 100%的正确划分.此外,由于现有黑名

单列表都是以二级域名或者三级域名的形式进行罗列,按照二级域名进行聚类,几乎所有(99.99%)域名分组都能得到 100%的正确划分.其次,聚类算法需要在保证分组正确性的前提下,尽可能多地减少分组数目.本文通过计算 1–|*G*|/|*D*|来衡量域名聚类算法的压缩率,其中,|*G*|为聚类后的分组数,而|*D*|为聚类前的域名数.如图 11 所示:按照二级域名的聚类算法有 97.4%左右的压缩率,依据连通二部图的域名聚类算法有 95.6%压缩率,而按照解析 IP 地址的域名聚类算法只有 86.4%的压缩率.

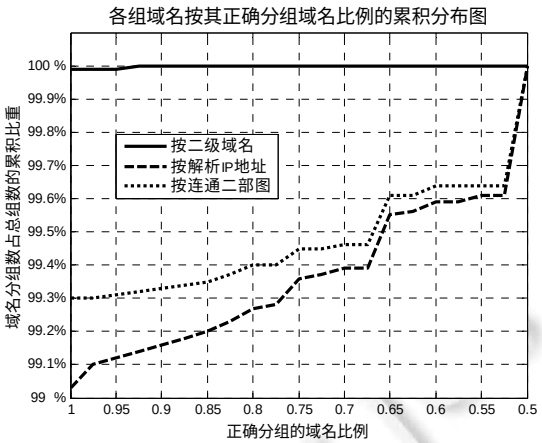


Fig.10 Cumulative distributionof domain name groups according to their correct clustering proportion

图 10 各组域名按其正确分组的域名比例的累积分布图

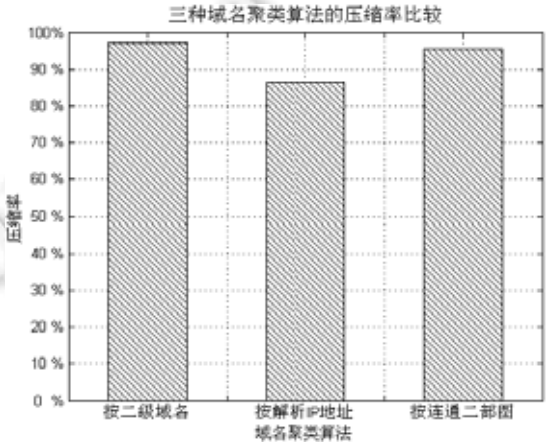


Fig.11 Compressibility comparison of the three domain name clustering algorithms

图 11 3 种域名聚类算法的压缩率比较

综合上述两方面评估结果,按照二级域名的聚类算法同时拥有最大正确分组比例和最大压缩比例.因此,本文选取该聚类算法,分别对标准域名集和实测域名集进行分组,以组为 DNS 活动监测的基本单元.

2.2 测度选择

为了观察域名的 DNS 活动规律,本文从依赖性和使用位置两个方面提出两组特征测度:前一组从用户角度观察域名的外在使用情况;后一组关注区域文件中记录的域名内部资源配置.虽然两组测度集彼此独立,但是数目较大.一方面会增加测度计算的复杂度,进而增大系统开销和时间开销;另一方面,冗余测度也会降低检测结果的准确率.为此,本文针对两组测度,分别运用测度选择算法 CFS[28]选取最优测度集.对于依赖性测度组,选取测度(1) *ipNum*,(2) *ipRatioPerUnit*,(3) *ipNumPerCity*,(4) *liveSpan*,(5) *clientAppearFreq*,(6) *clientLine Distance* 和(7) *abruptRatio*;而对于使用位置测度组,选取(8) *ripTimesDistribution*,(9) *ripSdomainTNum*,(10) *ripsdomainMNum*,(11) *ripDiffRatio*,(12) *nsDiffRatio* 和(13) *nsLiveMSpan*.此处需要说明的是:在这 13 个测度中,除测度(1)和测度(4)是已有测度外,其余 11 个测度都是改进测度或者新测度.针对单位数/城市数两个空间测度无法有效区分合法域名和垃圾邮件域名,测度(2)、测度(3)进一步测量空间密度;为了进一步观察长期活跃域名的活跃情况,测度(5)继续测量活跃频率;测度(6)可以缓解欧拉距离测量中,曲线绝对高度对周期性检测的影响;测度(7)通过平均值和最大值的简单比值运算,可以降低 CUSUM 等算法在曲线变点分析时的计算复杂度;最后,为了识别混入 FFSN 的 CDN,降低检测的误报率.本文通过分析两者间的内在差异,提出一组新的测度(测度(8)~测度(13)),用于评估承载服务的网络基础架构的可靠性.

2.3 多分类器检测算法

目前,关于分类算法的研究已经相当成熟.Statlog 项目针对卫星影像等 8 个实际分类问题,考察了 18 种有监督的机器学习算法的分类效果.研究结果表明,没有一个算法能够同时适用于上述所有类别的分类问题.本文并不研究具体算法的改进和优化,而是关注何种分类器或者分类器组合更适合本文的数据集.

为了检测恶意域名,本文从依赖性和使用位置两个方面提出两组特征测度.若使用单个分类器进行异常检

测,则需要同时统计和分析两组测度.考虑到高维数据的处理系统开销大,且两组测度之间彼此独立,本文提出多分类器检测模型 DAOS:针对两组特征测度,各自选用准确率高且时间开销低的分类器进行检测,而后,通过加权平均的数学统计方法对上述两方面的检测结果进行数据融合,得到最终结果.实验观察发现,C4.5 分类器对两组测度都拥有最大的检测准确率、最小的假阳性和假阴性以及较小的时间开销.因此,选用 C4.5 分类器分别对两组测度进行 DNS 活动异常检测.为了后续叙述的方便,不妨假设对于任意一个域名对象 d_l,C4.5 分类器使用依赖性测度组将 d_l 检测为恶意域名的置信度为 $P_{dep}(d_l,C_{Bad})$,而使用使用位置测度组将 d_l 检测为恶意域名的置信度为 $P_{loc}(d_l,C_{Bad})$.

$$p(d_l,C_{Bad})=\alpha*P_{dep}(d_l,C_{Bad})+(1-\alpha)*P_{loc}(d_l,C_{Bad}) \tag{25}$$

$$\psi=\sum_{d_l\in D}(p'(d_l,C_{Bad})-p(d_l,C_{Bad}))^2 \tag{26}$$

为了合并两方面检测结果,采用加权平均的数学统计方法进行融合(见公式(25)),其中,参数α和 $1-\alpha$分别表示两组测度检测结果的权重,且$\alpha\in[0,1]$.为了估计参数α,本文在标准样本集的基础上,采用最小二乘法(见公式(26)),其中,$P'(d_l,C_{Bad})$表示各域名样本的实际置信度:若 d_l 是恶意域名,则值为 1;否则为 0.当α=0.55 时,ψ值最小,此时,多分类器算法 DAOS 的检测准确率为 93.9%,假阳性为 1.3%,假阴性为 4.8%,时间开销为 1.94s.如图 12 和图 13 所示:在单分类器算法中,C4.5 分类器具有最高的检测精度(准确率 93.7%,假阳性 1.7%,假阴性 4.6%)和相对较小时间开销(1.98s).与之相比,本文提出的多分类器算法提高 0.2%的准确度,降低 0.4%的假阳性和 0.4s(相对降低 20.2%)的时间开销,但假阴性增加 0.2%.

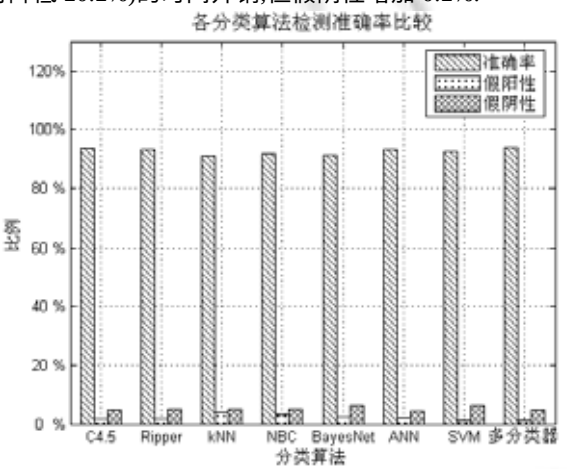


Fig.12 Accuracy comparison of these classification algorithms

图 12 各分类算法检测准确率比较

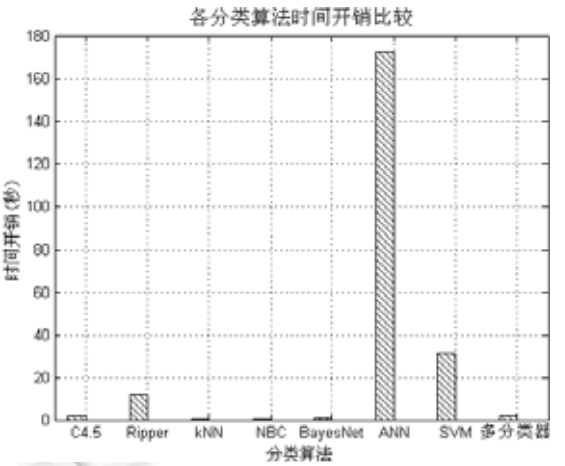


Fig.13 Time overhead comparison of these classification algorithms

图 13 各分类算法时间开销比较

3 实验结果

3.1 数据集

(1) 实测数据集

主干网 DNS 流量 *JSERNET_DNS*:2015 年 3 月 1 日~5 月 31 日,在中国教育科研网江苏省网边界一个接入点路由器上,连续 3 个月采集流经的 DNS 交互报文.

主干网流测量数据 *JSERNET_NETFLOW*:同样,从 2015 年 3 月 1 日~5 月 31 日,在中国教育科研网江苏省网边界同一个接入点路由器上,连续 3 个月采集流经的 NetFlow 流数据.

(2) 标准域名集

合法域名集 *Good_Domain_Set*:考虑到越普及通用的域名,其为合法域名的可能性越大.本文选取 Alex[22]网

站连续 3 次排名前 1 万且在 *JSERNET_DNS* 中出现过的域名,同时去除曾经在黑名单中出现过的域名.合法域名集一共包含 5 438 个二级域名对象,其中包含 161 个 CDN 二级域名对象.

恶意域名集 *Malicious_Domain_Set*:选取僵尸网络[23,24]、钓鱼网站[25]、垃圾邮件[26]和恶意软件[23,24]等黑名单中出现过、并且在 *JSERNET_DNS* 中出现过的域名.为保证恶意样本集的干净,进一步删除 Alex 网站 3 次排名前 100 万的域名.恶意域名集一共包含 5 533 个二级域名对象,其中,僵尸网络二级域名对象 2 846 个、钓鱼网站二级域名对象 663 个、垃圾邮件二级域名对象 937 个、其他恶意软件二级域名对象 1 087 个.

3.2 准确率

对于 DNS 流量检测算法而言,设计和实现的关键在于检测所依据的测度和采用的分类算法.目前,关于分类算法的研究无论是有监督的机器学习算法还是无监督的聚类或者分类算法都相当成熟.因而,测度的选取成为当前各 DNS 流量检测算法关注的重点.表 1 中,与本文 DAOS 系统在检测目标、数据源和实验环境上相似的工作有 Notos[19],Exposure[20,21]和 Kopis[15],它们都是针对上层 DNS 流量的通用域名检测算法,都使用了域名访问活动特征和资源记录特征,此外,3 项先前的工作还增加了域名字面特征和黑名单证据特征.

Table 1 The list of DNS traffic detection algorithms

表 1 DNS 流量检测算法列表

项目/人名	测度集(括号中的数字代表具体测度个数)	检测算法
Notos[19]	①域名字面特征(17);②资源记录 *A* 特征(18);③黑名单证据(6).	有监督机器学习
Exposure[20,21]	①域名字面特征(2);②资源记录 *A* 及 *TTL* 特征(17);③DNS 查询时间分布(7).	有监督机器学习(C4.5)
Kopis[15]	①DNS 查询者 RDNS 的空间分布及用户规模(13+10);②黑名单证据(9).	有监督机器学习(随机森林)
DAOS(本文)	①依赖性,即,DNS 查询用户的空间、时间分布特征(7); ②使用位置,即,资源记录 *A*,*NS* 及 *TTL* 特征(6).	有监督机器学习(多分类器)

DAOS 从依赖性和使用位置两个方面,一共提出 13 个测度.为了证明该测度集的重要性,本文基于相同的标准数据集和分类算法,分别选用 4 项工作中不同的测度集 C_1~C_4,比较它们的检测准确率,具体实验过程如下:首先,选取标准域名集 *Good_Domain_Set* 和 *Malicious_Domain_Set* 中标记过合法/恶意的二级域名作为观测对象集 O;其次,基于 DNS 数据 *JSERNET_DNS* 和 NetFlow 数据 *JSERNET_NETFLOW*,根据文献中测度集 C_i 的具体定义,统计每个二级域名对象的测度组;而后,统一使用单个 C4.5 分类器对标记过的二级域名对象集进行分类,有监督的机器学习方法需要提供训练集和测试集,本文采用十字交叉验证法,将域名对象集 O 划分成 10 份,每次选 9 份训练 1 份测试,最终得到的检测结果为这 10 次的平均结果;最后,从检测准确率、假阳性和假阴性 3 个方面比较 4 个测度集相应的检测结果.这里有两点需要说明:(1) 只关注域名访问活动特征和资源记录特征两方面的测度,由于域名字面特征检测精度不高,证据特征过多依赖外界黑名单的正确性,DAOS 现阶段并未考虑这两方面的特征测度,但是下阶段仍然可以增加它们作为辅助测度,因此算法比较时,出于公平性的考虑,Notos,Exposure 和 Kopis 这 3 项工作也不计算域名字面特征和黑名单证据特征;(2) 多分类器虽然在时间开销上优于单分类器(降低 20%),但是在检测准确率上相差不大(准确率只有 0.2%的增加),因此,该节对于所有测度集都使用单个 C4.5 分类器进行分类.

如图 14 所示,DAOS 虽然只使用了 13 个测度,却具有最高的检测精度(准确率 93.7%,假阳性 1.7%,假阴性 4.6%).为了探查其原因,本文基于表 1 给出的测度集进行分析.

- 首先,Notos 关注资源记录 *A* 的地理位置分布和逻辑归属分布,但是忽略了访问活动特征;而 Kopis 观测域名访问活动中,DNS 查询者“RDNS”的空间分布和用户规模,却未使用资源记录特征.与之相比,DAOS 结合了域名的访问活动特征和资源记录特征,因而检测精度具有较大幅度的提升:比 Notos 准确率提高 9.5%,假阳性降低 7.4%,假阴性降低 2.1%;比 Kopis 准确率提高 8.3%,假阳性减少 5.6%,假阴性也降低 2.7%;
- 其次,Exposure 虽然也综合了访问活动特征和资源记录特征,但是只观察资源记录 *A* 以及 *TTL* 的空间分布特征,只测量访问活动中 DNS 查询的时间分布特征;而 DAOS 全面分析了资源记录 *A*,*NS* 以及 *TTL*

的时间和空间分布特征,并借助流数据间接测量用户访问活动的时间和空间分布特征;

- 另外,实际观察发现:平均每个域名的查询用户数有 340 个,而 RDNS 数只有 210 个.即,从用户视角可以比 RDNS 视角更好地观测访问活动特征.因此,DAOS 比 Exposure 拥有更高的检测精度(准确率上升 2.8%,假阳性减少 2.8%,假阴性不变);
- 最后,值得说明的是:Notos 使用了 18 个测度,Exposure 使用了 24 个测度,Kopis 使用了 23 个测度;而本文通过使用测度选择算法 CFS,只选取了 13 个测度,大概只有它们的一半.

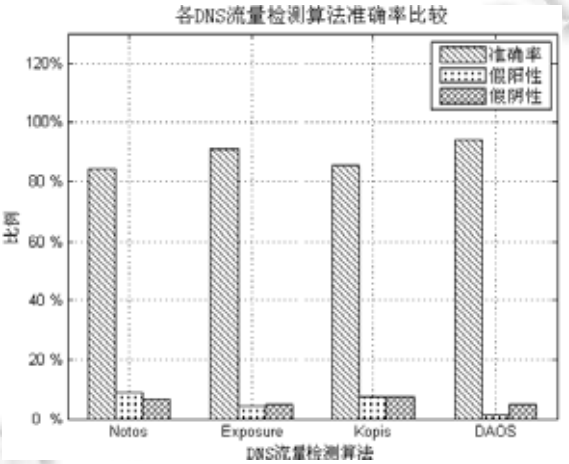


Fig.14 Accuracy comparison of these DNS traffic detection algorithms

图 14 各 DNS 流量检测算法准确率比较

实验结果表明:本文从依赖性和使用位置两个方面提出的两组特征测度,要优于现有检测算法所使用的域名访问活动特征测度和资源记录特征测度.因而,可以改进和补充现有检测算法中的测度,提高主干网 DNS 流量实时检测的精度.

3.3 影响因子

恶意域名的识别,需要事先观察和统计该域名对象的 DNS 活动特征.一般而言,观测的时间越长,检测的准确率越高;观测的时间越短,检测的准确率也会相对下降.即,准确率和观测时间长度此消彼长.如何根据用户对检测实时性的需求选择两者间合适的平衡点,是需要研究的一个重点.

本节重复第 3.2 节中的实验过程,保持数据集、测度集(13 个测度)和多分类算法不变,调节测度统计的时间窗口长度(2 小时、4 小时、8 小时、16 小时、1 天、2 天、1 星期),即,每个域名对象需要持续观测的时间长度.而后,同样使用十字交叉验证法,从准确率、假阳性和假阴性这 3 个方面评估算法的检测精度.如图 15 所示:随着观测时间窗口长度的增加,算法的检测准确率确实有所增加,但增幅不大;与此同时,假阳性和假阴性也稍有减少.

当观测时间长度从 2 小时增加到 1 星期后,检测准确率从 90.5%上升到 93.9%(增加 3.4%),假阳性从 2.9%下降到 1.3%(减少 1.6%),假阴性也从 6.6%下降到 4.8%(减少 1.8%).结果表明:DAOS 具有较高的实时检测能力,经过 2 个小时的实时监测,就能达到 90.5%的准确率.若用户期望获得较高的准确率,则需要适当延长域名观测的时间长度.

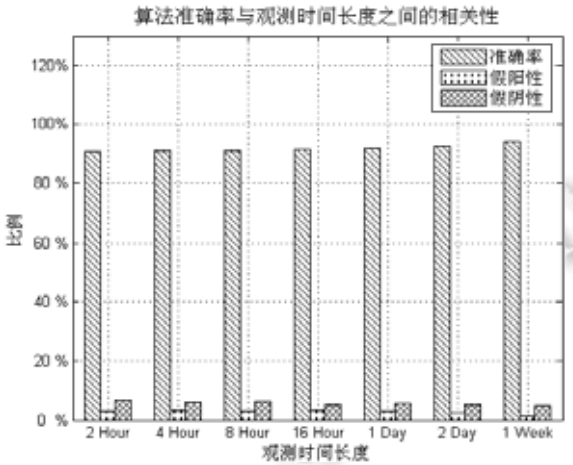


Fig.15 Correlation between theobservation intervaland the algorithm's detection accuracy

图 15 算法准确率与观测时间长度之间的相关性

3.4 实用性测试

为了验证 DAOS 检测算法的实用性,本文在中国教育科研网江苏省网边界的一个接入点路由器上,实时采集流经的 DNS 流量和 NetFlow 流记录数据,同时转发给后端 DAOS 检测服务器.DAOS 系统运行前,设定时间窗口长度为 2 小时,同时选用第 3.1 节中的两个标准域名列表 *Good_Domain_Set* 和 *Malicious_Domain_Set* 作为最初的训练样本集.DAOS 实际运行过程中包含 3 个重要环节:首先,当标准域名列表发生变更时,系统在接下来的一个时间片内,需要重新统计标准域名列表中所有标记域名的测度值,生成新的训练样本集;其次,DAOS 维护着一张二级域名列表,当出现新的二级域名对象时,才统计其测度值,并使用多分类器和训练样本集进行分类;最后,系统定期从检测结果中,选择置信度较大的二级域名对象,更新标准域名列表,实现 DAOS 算法的自学习.

从 2015 年 6 月 1 日~6 月 30 日,平均每天新出现的二级域名对象有 2.89 万个,而 DAOS 每天检测到的可疑二级域名对象有 2 852 个,一个月共发现 8.57 万个可疑域名对象.为了验证检测结果的准确率,采用抽样检测的方法,通过统计抽样样本的准确率来评估整个检测结果集的准确率.本文选用千分之一的抽样比例,从可疑域名对象集中随机抽出 857 个二级域名样本,通过查询在线黑名单[23−26]和可疑文件分析服务网站 VirusTotal[29],或者手工验证的方法确定样本类别.其中,出现在钓鱼网站[23]中的域名有 1 个,出现在恶意域名黑名单[23,24]中的域名有 3 个,DGA 域名有 22 个.VirusTotal 确定为恶意的域名有 31 个,安全的域名有 53 个.另外,有 179 个域名未经注册,有 137 个域名所辖网站包含色情、赌博和恶意销售等内容,75 个域名所辖网站无效、过期或者正在维护中,224 个域名在一个月内的活跃时间长度不足 1 小时,还有 132 个域名无法进行确认.

此外,针对上述 75 个无效、过期和正在维护中的域名网站进行追踪分析:首先,这 75 个域名网站在原始 DNS 数据中存在正确的解析响应报文,能够提取出域名映射的 IP 地址;同时,在流测量数据中也可以找到这些 IP 地址的通信会话信息,说明它们当时都活跃着.但是现在进行手工检测发现,这些域名网站都不再提供正常服务.其中,2 个网站长期处于维护中,47 个域名已经过期超过 3 个月;剩余 26 个网站无响应.通常情况下,一个合法网站是不会停止服务、不续费域名以及长期进行维护的.为了验证这一假设,本文关注 2015 年 6 月 Alex 网站排名前 2 万名中的 1 万个网站(绕过知名网站),发现其中 96.5%的网站在 2016 年 9 月还活跃着.也就是说,这 75 个域名网站如果是合法网站的话,则 1 年后至少应该有 72 个网站能够正常提供服务.另一方面,恶意网站存活时间都不长.如图 5 所示,生命周期长度超过 85 天的恶意域名不超过 5%.换句话说,这 75 个域名网站如果是恶意网站的话,则 1 年后最多只有 4 个网站还能够正常提供服务.最后,综合两方面的假设分析可知,这 75 个域名网站是恶意网站,置信度超过 95%.

综上所述,若除去无法确认的 132 个域名,则剩余 725 个域名.其中,能够确认的安全域名有 53 个,即误报率至少为 7.3%.而能够确认的恶意域名有 269 个(包括 4 个黑名单域名,22 个 DGA 域名,31 个 VirusTotal 确认的恶

意域名,137 个色情、赌博和恶意销售网站,75 个无效、过期和正在维护中的网站).考虑到合法域名的生命周期长度基本上都超过 1 天,且正常情况下,未经注册的域名是不应该出现在 DNS 报文中,若把 224 个活跃时间长度不足 1 小时的域名以及 179 个未经注册的域名也看成恶意域名,则 DAOS 检测准确率可以达到 92.7%.这与第 3.3 节中统计时间窗口设定为 2 小时的准确率 90.5%和误报率 2.9%相符,说明 DAOS 在实际运行时也能保证较高的实时性和检测精度.

4 总 结

为了保障 ISP 主干网运行安全,本文实时检测流经主干网边界的 DNS 交互报文,研究适用于主干网环境的上层 DNS 流量检测算法.

首先,借助主干网边界采集的流测量数据,间接获取域名与终端用户间的相互关系,来弥补 DNS 缓存机制屏蔽终端用户解析请求的不足;并在此基础上提出域名依赖性的概念,通过测量用户的空间规模和活跃度,从用户角度观察域名的外在使用情况;而后,为了提高算法的检测精度,观察资源记录 *A* 和 *NS* 中隐藏的域名使用位置信息,从域名自身角度关注其内部资源部署情况.在传统测量 IP-Flux 和 NS-Flux 特征的基础上,通过分析 CND 和 FFSN 内在的差异性,提出一组新的测度以识别混入 FFSN 中的 CDN;再者,主干网 DNS 流量检测需要对海量的域名对象进行实时或准实时地 DPI 检测.为了提高算法的检测效率,一方面按照相同的二级域名将域名划分成组,以组为单位进行检测;另一方面使用 CFS 算法选择最优测度集,并基于两个测度集提出多分类器的检测模型;最后,为了兼顾检测算法的及时性和准确率,同时关注域名的当前活动特征和相关对象的历史活动行为,使用较长时间的历史相关数据来弥补当前时间窗口长度较短导致数据不足的问题.

实验观察发现:对于某个域名对象,在不依赖于先验知识的前提下,经过两个小时的 DNS 活动监测,检测准确率可以达到 90.5%,假阳性和假阴性分别为 2.9%和 6.6%.若用户期望获得更高的检测精度,持续监测一周,准确率可以提升到 93.9%,假阳性和假阴性也能减少到 1.3%和 4.8%.同时,观察发现:在不依赖域名字面特征和黑名单先验知识的前提下,本文的 DAOS 算法与现有面向上层 DNS 流量的通用域名检测算法 Notos,Exposure 和 Kopis 相比,虽然使用最小数目的测度集,但是具有最高的检测精度.最后,在实用性测试中,当统计时间窗口设定为 2 小时时,DAOS 也能获得 92.7%的准确率和 7.3%的误报率.

综上所述,本文提出的依赖性和使用位置测度组以及多分类器模型,可以有效地用于主干网环境下的实时 DNS 流量监测;也可以作为现有算法的补充,提高它们的检测精度.

References:

[1] Levchenko K, Pitsillidis A, Chachra N, Enright B. Click trajectories: End-to-End analysis of the spam value chain. In: Butler K, ed. Proc. of the IEEE Symp. on Security and Privacy. IEEE, 2011. 431−446. [doi: 10.1109/SP.2011.24]

[2] Bot traffic report. 2015. https://www.incapsula.com/blog/bot-traffic-report-2015.html

[3] APWG phishing activity trends report. 2016. http://www.antiphishing.org/resources/apwg-reports/

[4] Schonewille A, Helmond DJV. The domain name service as an IDS [MS. Thesis]. University of Amsterdam, 2006. http://staff.science.uva.nl/~delaat/snb-2005-2006/p12/report.pdf

[5] Chatzis N, Popescu-Zeletin R. Flow level data mining of DNS query streams for email worm detection. In: Corchado E, Zunino R, Gastaldo P, Herrero A, eds. Proc. of the Int'l Workshop on Computational Intelligence in Security for Information Systems (CISIS 2008). Berlin, Heidelberg: Springer-Verlag, 2009. 186−194. [doi: 10.1007/978-3-540-88181-0_24]

[6] Chatzis N, Popescu-Zeletin R. Detection of email worm-infected machines on the local name servers using time series analysis. Journal of Information Assurance and Security, 2009,4(3):292−300.

[7] Chatzis N, Popescu-Zeletin R, Brownlee N. Email worm detection by wavelet analysis of DNS query streams. In: Dasgupta D, Zhan J, eds. Proc. of the IEEE Symp. on Computational Intelligence in Cyber Security (CICS 2009). Nashville: IEEE, 2009. 53−60. [doi: 10.1109/CICYBS.2009.4925090]

[8] Chatzis N, Brownlee N. Similarity search over DNS query streams for email worm detection. In: Awan I, ed. Proc. of the 2009 Int'l Conf. on Advanced Information Networking and Applications (AINA 2009). Bradford: IEEE, 2009. 588−595. [doi: 10.1109/AINA.2009.132]

[9] Choi H, Lee H, Kim H. Botnet detection by monitoring group activities in DNS traffic. In: Wei D, ed. Proc. of the 7th IEEE Int'l Conf. on Computer and Information Technology (CIT 2007). Fukushima: IEEE, 2007. 715−720. [doi: 10.1109/CIT.2007.90]

[10] Choi H, Lee H, Kim H. BotGAD: Detecting botnets by capturing group activities in network traffic. In: Bosch J, Clarke S, eds. Proc. of the 4th Int'l ICST Conf. on Communication System Software and Middleware (COMSWARE 2009). Dublin: ACM Press, 2009. 2–2. [doi: 10.1145/1621890.1621893]

[11] Jiang N, Cao J, Jin Y, Li LE, Zhang ZL. Identifying suspicious activities through DNS failure graph analysis. In: Gunes MH, ed. Proc. of the 18th IEEE Int'l Conf. on Network Protocols (ICNP 2010). Kyoto: IEEE, 2010. 144–153. [doi: 10.1109/ICNP.2010.5762763]

[12] Yadav S, Reddy ALN. Winning with DNS failures: Strategies for faster botnet detection. In: Rajarajan M, Piper F, eds. Proc. of the Security and Privacy in Communication Networks. London: Springer-Verlag, 2012. 446–459. [doi: 10.1007/978-3-642-31909-9_26]

[13] Lee J, Kwon J, Shin HJ, Lee H. Tracking multiple C&C botnets by analyzing DNS traffic. In: Fahmy S, ed. Proc. of the 6th IEEE Workshop on Secure Network Protocols (NPSec 2010). Kyoto: IEEE, 2010. 67–72. [doi: 10.1109/NPSEC.2010.5634445]

[14] Lee J, Lee H. GMAD: Graph-Based malware activity detection by DNS traffic analysis. Journal Computer Communications, 2014, 49(12):33–47. [doi: 10.1016/j.comcom.2014.04.013]

[15] Antonakakis M, Perdisci R, Lee W, Li NV, Dagon D. Detecting malware domains at the upper DNS hierarchy. In: Wagner D, ed. Proc. of the 20th USENIX Conf. on Security (SEC 2011). San Francisco: USENIX, 2011. 27–27.

[16] Thomas M, Mohaisen A. Kindred domains: Detecting and clustering botnet domains using DNS traffic. In: Chung CW, eds. Proc. of the 23rd Int'l Conf. on World Wide Web. New York: ACM Press, 2014. 707–712. [doi: 10.1145/2567948.2579359]

[17] Holz T, Gorecki C, Rieck K, Freiling FC. Measuring and detecting fast-flux service networks. Network and Distributed System Security Symp., 2008,1(5):487–492.

[18] Caglayan A, Toothaker M, Drapeau D, Burke D, Eaton G. Real-Time detection of fast flux service networks. In: Walter E, ed. Proc. of the 2009 Cybersecurity Applications & Technology Conf. for Homeland Security (CATCH 2009). Washington: IEEE, 2009. 285–292. [doi: 10.1109/CATCH.2009.44]

[19] Antonakakis M, Perdisci R, Dagon D, Lee W, Feamster N. Building a dynamic reputation system for DNS. In: Goldberg I, ed. Proc. of the 19th USENIX Conf. on Security (SEC 2010). Berkeley: USENIX, 2010. 18–18.

[20] Bilge L, Kirda E, Kruegel C, Balduzzi M. Exposure: Finding malicious domains using passive DNS analysis. In: Nishide T, ed. Proc. of the 18th Annual Network and Distributed System Security Symp. (NDSS 2011). Virginia: Internet Society, 2011. 195–211.

[21] Bilge L, Sen S, Balzarotti D, Kirda E, Kruegel C. Exposure: A passive DNS analysis service to detect and report malicious domains. ACM Trans. on Information and System Security (TISSEC), 2014,16(4):14–14. [doi: 10.1145/2584679]

[22] Alexa. 2015. http://www.alexa.com/topsites/

[23] DNS-BH malware domain blocklist. 2015. http://www.malwaredomains.com

[24] Malware domain list. 2015. http://www.malwaredomainlist.com

[25] PhishTank. 2015. http://www.phishtank.com

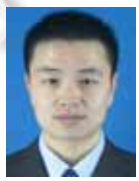
[26] Blacklist provided by joewein.net (JWSDB). 2015. http://joewein.net/spam/blacklist.htm

[27] Krishnamurthy B, Wills C, Zhang Y. On the use and performance of content distribution networks. In: Paxson V, ed. Proc. of the 1st ACM SIGCOMM Workshop on Internet Measurement (IMW 2001). New York: ACM Press, 2001. 169–182. [doi: 10.1145/505202.505224]

[28] Hall MA. Correlation-Based feature subset selection for machine learning [Ph.D. Thesis]. Hamilton: University of Waikato, 1999.

[29] VirusTotal. 2016. https://www.virustotal.com

张维维(1984－),男,江苏南通人,博士生,主要研究领域为网络安全.



刘尚东(1979－),男,博士生,CCF 专业会员,主要研究领域为网络安全.

龚俭(1957－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为网络管理,网络安全.

胡晓艳(1985－),女,博士,讲师,CCF 专业会员,主要研究领域为网络管理,下一代互联网.