基于解析数据的 DNS 安全评估与增强

毕设答辩

无 41 芦迪

指导老师：李星

Department of Electronic Engineering,
Tsinghua University

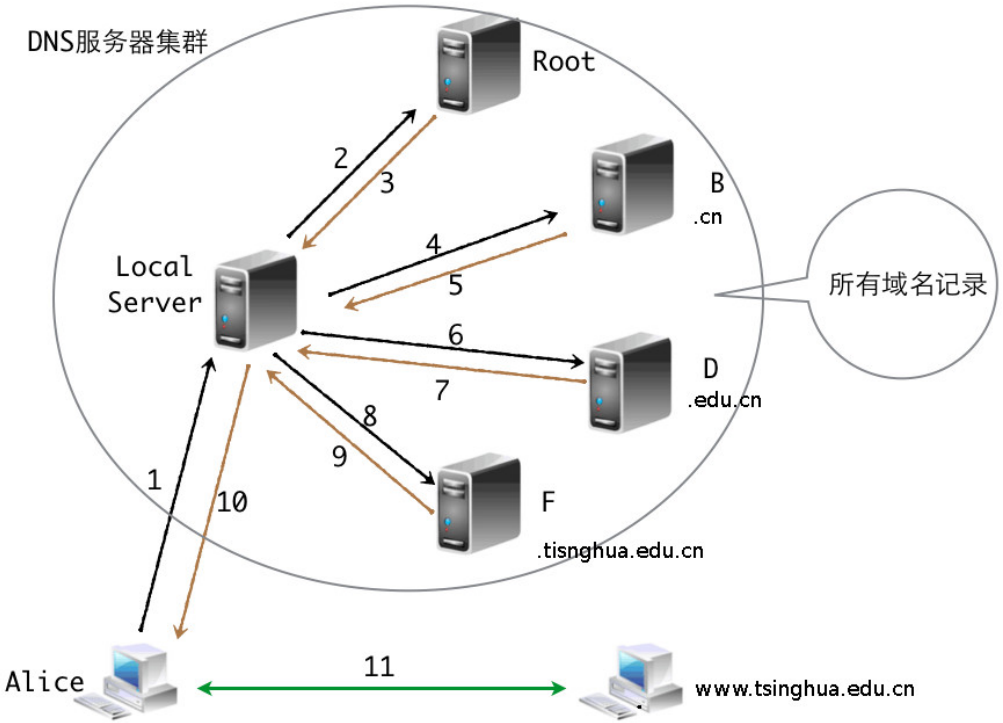
June 15, 2018

目录

目录

DNS: Domain Name System，域名系统

- 域名 (www.tsinghua.edu.cn) $\Longleftrightarrow$ IP 地址 (166.111.4.100)
- 重要的互联网基础设施，遭到攻击损失无法估量

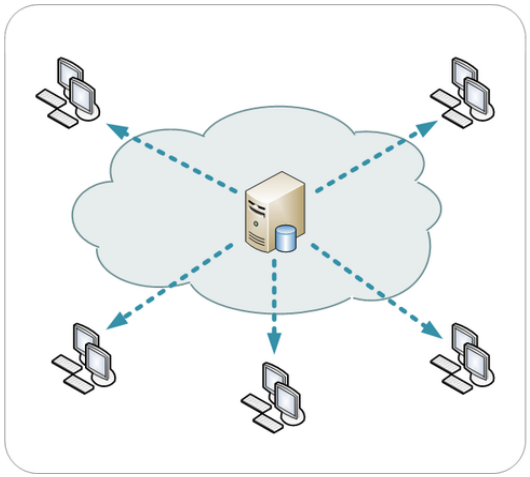


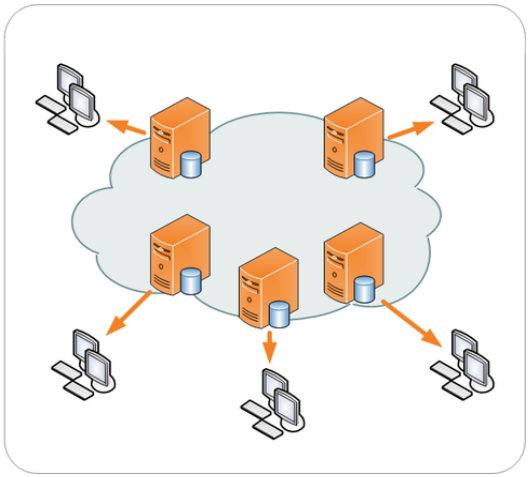
DNS 攻击

- DNS 易受攻击
  - 协议脆弱性：UDP、明文
  - 系统脆弱性：Open Resolvers、路径复杂
  - 常见攻击类型
    - DNS 劫持：将 DNS 请求定向到非法解析器
    - DNS 缓存中毒：将虚假域名数据注入到缓存中
  - DNS 攻击实例：
    - 2014 年 1 月 21 日下午，大陆境内发生了严重的 DNS 故障，所有的通用顶级域（.com/.net/.org 等）遭到 DNS 劫持/污染，所有域名被指向到一个位于美国的 IP 地址（65.49.2.178）
    - 2016 年 10 月 21 日，网络提供商 Dynamic Network Service 的域名服务器遭遇 DDoS 攻击，美国一个很大区域内的互联网在十个小时内无法访问 Twitter、Ebay、Netflix、Amazon、Paypal 等网站
    - .....

CDN 与负载均衡

- 因为 CDN 与负载均衡的普遍存在，域名解析情况更加复杂
  - CDN：Content Delivery Network
    - 部署边缘服务器，使用户就近获取所需内容，降低网络拥塞





  - 负载均衡：在多个服务器中分配负载，最优化资源使用，避免过载
- 一个域名会解析出多个 IP，一个 IP 也可以对应多个域名

网络安全增强

互联网天然缺乏安全性，通过完善协议可以增强安全性

- DNSSEC：DNS 安全扩展
  - 使用公钥密码机制，以 DNS 资源记录计算数字签名，验证数字签名确保解析结果真实
  - 提供数据来源、数据完整性和否定存在验证
- IPv6：下一代互联网协议
  - IPv6 内嵌 IPSec 协议族的 AH 和 ESP，从协议上提升安全性
  - IPv6 地址空间大，增大扫描难度，可减缓现有攻击
- DNS over HTTPS
  - 使用 HTTPS 传输解析信息，增强了客户端和递归解析器之间的隐私和安全性

前人工作

目前在对 DNS 数据进行测量、分析的领域内已经有很多的研究成果。

- 文献 [1]* 通过对 DNS 解析结果进行过滤，并对 HTTP 请求返回的内容进行聚类，系统地分析了非合法的 DNS 响应
- 文献 [2]† 利用开放解析器通过采集 DNS 解析数据来识别 CDN 的部署与解析中存在的网络干扰
- 文献 [3]‡采用一致性和独立的可验证性指标，通过对解析结果 IP 地址、自治域、HTTP 内容、HTTPS 证书等内容进行验证，实现了对网络审查的检测
- 还有其他文献对测量什么、如何测量、确定异常行为等方面进行了深入的研究

*Going Wild: Large-Scale Classification of Open DNS Resolvers

†Joint Analysis of CDNs and Network-Level Interference using Satellite

‡Global Measurement of DNS Manipulation

毕设内容

- 对权威服务器记录内容进行考察
  - 分析域名解析数据，了解域名或解析器对 IPv6、DNSSEC、HTTPS 的支持情况，评估安全现状
    - 根据 A 记录与 AAAA 记录，判断 IPv6 支持情况
    - 根据 RRSIG 记录，判断 DNSSEC 支持情况
    - 测试 (域名, IP 地址)，判断 HTTPS 支持情况
- 对开放递归服务器进行考察
  - Open Resolvers 基本情况
  - 对 DNS 解析返回的 IP 地址进行判别与分类
    - 结果正常：CDN？负载均衡？
    - 结果有误：DNS 劫持？缓存中毒？

目录

系统设计

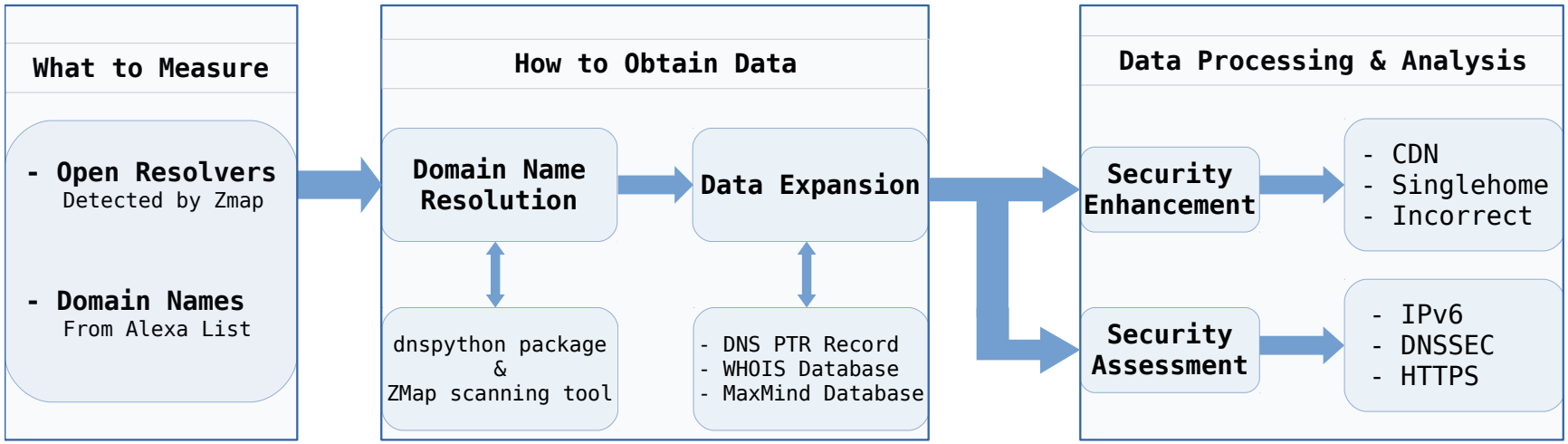


Figure: 整体系统设计

测量什么

- 对开放解析器状况进行评估：ZMAP 扫描 IPv4 空间，设置过滤条件，获得包含 1.6 万开放解析器的集合
  - 设置过滤条件：响应时间在 3 秒内、PTR 反向查询可获得结果.....
  - 使用 ZMap 进行扫描
    - 对 IPv4 网络空间进行快速测量的工具
    - 对单端口大规模测量有着很好的支持
    - 不记录状态，可在 45min 扫描整个 IPv4 空间
    - 支持 TCP SYN 、UDP 等多种扫描模式

测量什么

- 对开放解析器状况进行评估：ZMAP 扫描 IPv4 空间，设置过滤条件，获得包含 1.6 万开放解析器的集合
- 对域名安全状况进行评估：选择 Alexa 统计的全球访问量前 10000 的网站 $\times$ 手动选择的包含 40 个开放解析器的集合

Num.	Domain Name
1	google.com
2	youtube.com
3	facebook.com
4	baidu.com
5	wikipedia.org
6	yahoo.com
7	reddit.com
8	google.co.in
.....	.....

Table: Top Domain Names

Resolver	Owner	DNSSEC
8.8.8.8	Google	Y
4.2.2.2	MicroSoft	Y
9.9.9.9	IBM	Y
8.26.56.26	Comodo	Y
180.76.76.76	Baidu	Y
202.141.162.123	USTC	Y
223.6.6.6	Ali	N
114.114.114.114	114DNS	N
.....	.....	.....

Table: Open Resolvers

测量什么

- 对开放解析器状况进行评估：ZMAP 扫描 IPv4 空间，设置过滤条件，获得开放解析器的集合
- 对域名安全状况进行评估：选择 Alexa 统计的全球访问量前 10000 的网站 $\times$ 手动选择的包含少量开放解析器的集合
- 对 DNS 查询结果进行分类:Alexa Top 1000 $\times$ 扫描获得的开放解析器列表

获取数据

- dnspython 进行多样化查询
  - Python 的一个强大的 DNS 工具包
    - 采用多种模式进行 DNS 查询的操作
    - 对收到的 DNS 响应进行解析
  - DNSSEC 查询、A 记录查询、AAAA 记录查询、基于 tcp 的查询
  - 单线程因网络延迟导致低效率 $\Rightarrow$ 多线程，维护查询线程池
- ZMap 进行快速 DNS 查询
  - 可以快速进行大量 DNS 查询
  - 以 Open Resolvers 为白名单，构建 DNS 请求包并发送
  - 解析回包并保存

例：

```
zmap -M udp -p 53 –probe-args=file:tsinghua.edu.cn.pkt -N 100
```

数据扩充

采用多种方式对数据进行扩充，便于之后的处理：

- DNS PTR Record
  - PTR 记录包含在“in-addr.arpa”这一域名下
  - 通常由控制这个 IP 的组织进行维护
  - 对属于已知服务的 IP 通常提供规范的名称
- WHOIS Database
  - 包含 IP 地址的所有权信息
- MaxMind Database
  - 获取 IP 地址的地理位置信息
- BGP 路由表数据
  - IP 地址的 AS 编号

安全评估

- 通过 DNS 查询可以了解 IPv6、DNSSEC 等协议部署情况
  1. 根据 A 记录与 AAAA 记录，判断 IPv6 支持情况

```
;; ANSWER SECTION:
xinhuanet.com.          3600    IN      A       202.108.119.194
xinhuanet.com.          3600    IN      A       202.108.119.193

;; AUTHORITY SECTION:
xinhuanet.com.          280     IN      NS      ns2.cdns.cn.
xinhuanet.com.          280     IN      NS      ns3.cdns.cn.
xinhuanet.com.          280     IN      NS      ns1.cdns.cn.

;; ADDITIONAL SECTION:
ns1.cdns.cn.            86      IN      A       125.208.45.1
ns2.cdns.cn.            393     IN      A       125.208.46.1
ns3.cdns.cn.            1382    IN      A       125.208.47.1
ns1.cdns.cn.            86      IN      AAAA    2001:dc7:ffeb::1
ns2.cdns.cn.            393     IN      AAAA    2001:dc7:ffec::1
```

  2. 根据 RRSIG 记录，判断 DNSSEC 支持情况

```
;; ANSWER SECTION:
paypal.com.             124 IN A 64.4.250.33
paypal.com.             124 IN A 64.4.250.32
paypal.com.             124 IN RRSIG A 5 2 300 (
                               20180205064703 20180106054703 11811 paypal.com.
                               rXHdlxddoLYYUjTuDxRLmtMxkTNdHVKaYBGif8KC03Jp
                               AWjl564STz3OQcJSpyurpJJYXCHVwkUt/X6RUfkw3neQ
                               gH0CjspEEiVxaOwEqZNEcWZBUDm77Vy4t1BpYiogNWGR
                               rKqezt9lon48zCZj9Zi0gdAE4qYNjvi1Glp9HCA= )
```

安全评估

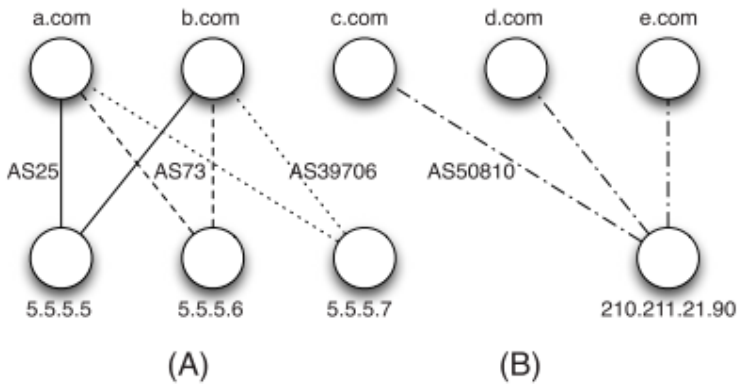
- HTTPS 部署情况及正确性
  - 使用数字证书来验证站点的身份，加密用户和站点之间的数据交换
  - 对于解析到的 IP 地址，对其 443 端口发起连接请求，验证是否支持 HTTPS
  - 对于支持 HTTPS 的 IP 地址，对其证书进行验证，将验证通过的 (domain, ip, resolver) 对标记为正确

安全增强

- 通过 DNS 解析，由域名得到 IP 地址，我们将其分为以下类别：
  1. 单归属域名返回唯一 IP 地址
  2. CDN 或负载均衡返回的结果
  3. DNS 劫持、缓存中毒或其他错误结果
- 对解析结果的识别流程
  1. IP 地址信息扩充，进行初步聚合
  2. 采用文献 [2] 中的联合分析算法，实现对 CDN 与网络干扰的分别聚类
  3. HTTPS 对解析到的 IP 地址进行验证，作为 Groundtruth

联合聚类算法

- 定义两个函数 DomainSimilarity 与 IPTrust
  - DomainSimilarity 表示两个域名的相似程度
  - IPTrust 表示 IP 是域名的正确解析结果的可信度
  - (A) 中，a.com 与 b.com 应当拥有一个高的 DomainSimilarity，因为它们被解析到相同的 IP 地址上
  - (B) 中，IP 地址 210.211.21.90 应当拥有一个低的 IPTrust，因为很多不相关的域名解析了它
  - 给定初始值，迭代计算，直到收敛



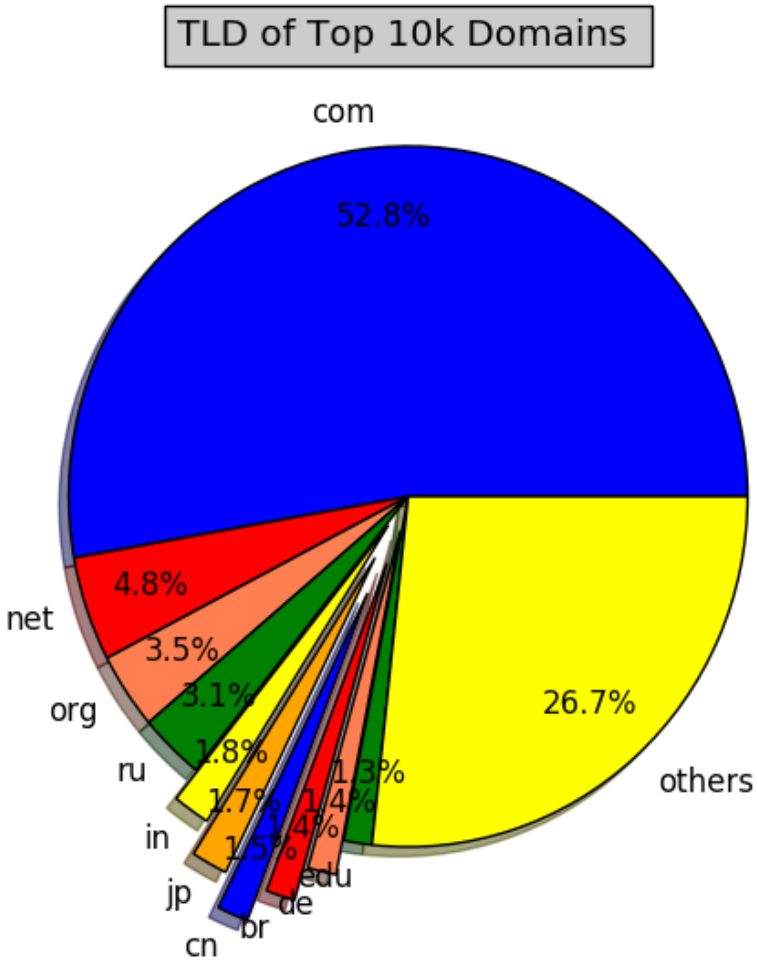
- 文献 [2] 中通过对比域名图标与 IP 地址图标对结果进行验证
- 我们通过检验 HTTPS 结果对结果进行验证

部署与实施

- 实验平台信息
  - 实验平台：DigitalOcean 服务器
  - 操作系统：Ubuntu 16.04
  - 公网 IP：206.189.75.45，2604:a880:2:d0::207d:a001
  - 数据保存：MySQL, phpmyadmin
- 数据的获取
  1. ZMap 对 IPv4 空间进行扫描并过滤，得到 16488 个解析器
  2. dnspython 进行 DNS 查询操作，Alexa Top 10000 域名在一个手动选取的 40 个解析器的列表上进行查询，共获得 365517 个查询结果，802058 个 (domain,res,ip) 三元组，74894 个 (domain,ip) 二元组
  3. ZMap 进行 DNS 查询操作，Alexa Top 1000 域名在 16488 个开放解析器上进行查询，最终获得 926 万个 (domain, res, ip) 三元组，16 万 (domain, ip) 二元组

目录

Alexa Top 10000 顶级域名分布情况：



IPv6、DNSSEC、HTTPS

IPv6、DNSSEC、HTTPS 整体及在 com、net 和 cn 顶级域下的支持情况：

	Count	HTTPS	IPv6	DNSSEC
ALL	8849	5376，60.8%	2159，24.4%	290，3.28%
com	4744	2874，60.6%	1073，22.6%	102，2.15%
net	409	239，58.4%	139，34.0%	14，3.42%
cn	156	42，26.9%	6，3.85%	1，0.64%

从表中可以看出：

- cn 顶级域名下 HTTPS、IPv6、DNSSEC 的支持情况都要远远低于平均
- net 顶级域名下 IPv6 和 DNSSEC 的支持情况要高于平均

联合分布情况

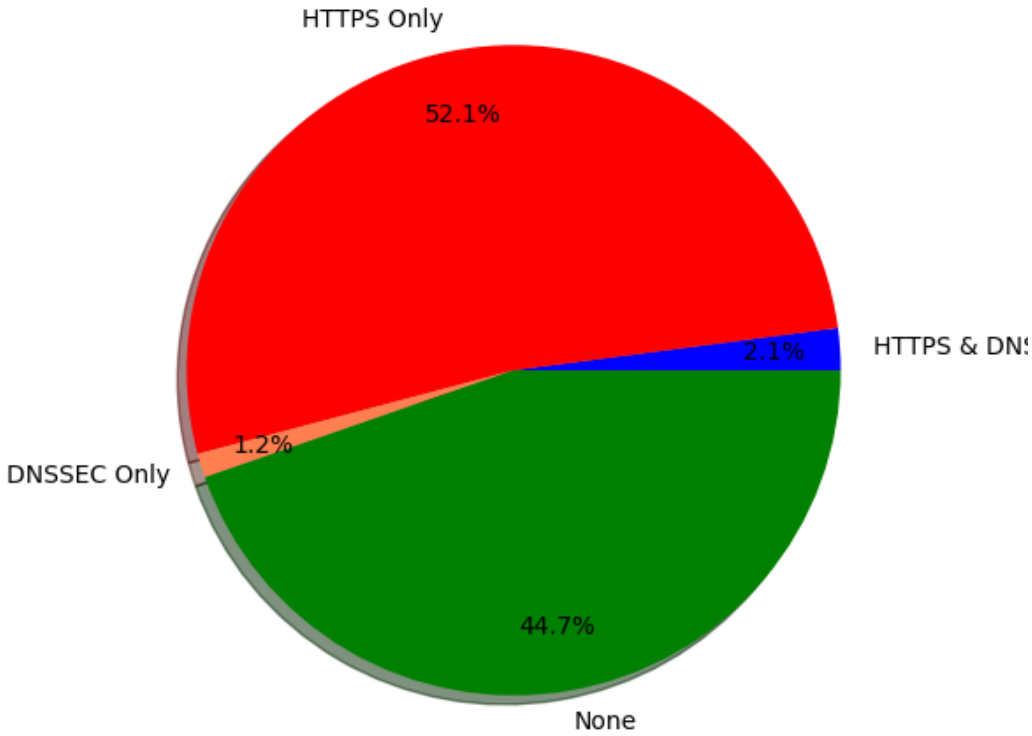
下面我们考察 HTTPS、IPv6、DNSSEC 在整体及不同顶级域名下的联合分布情况，如下表所示：

	HTTPS+IPv6	HTTPS+DNSSEC	IPv6+DNSSEC	IPv6+DNSSEC+HTTPS
ALL	1796	201	114	88
com	874	82	42	38
net	120	10	6	5
cn	2	0	0	0

经过进一步的分析，我们发现：

- 在部署 IPv6 同时，部署 HTTPS 的概率很大
- DNSSEC 的部署情况与 IPv6 和 HTTPS 的部署情况无关

HTTPS 与 DNSSEC

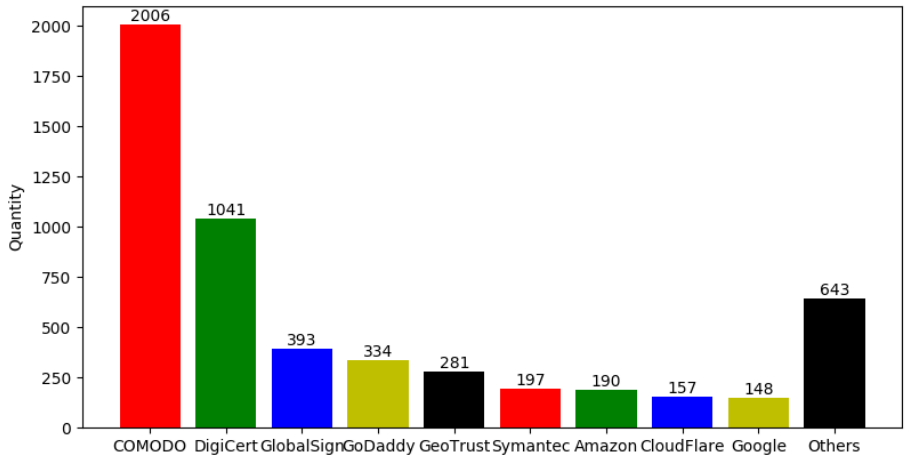


- 仅仅 HTTPS 验证成功或者两者都没有验证成功的结果是最常见的
- DNSSEC 验证通过而 HTTPS 没有完成验证的情况最令人费解
  - 考察了这个类别，其中有很多成人网站、政府机构和教育机构的网站

HTTPS 证书服务商

HTTPS 证书服务商统计：

Service Provider	Quantity	Percent
COMODO CA Limited	2006	37.22%
DigiCert Inc	1041	19.31%
GlobalSign	393	7.29%
GoDaddy.com, Inc.	334	6.19%
GeoTrust Inc.	281	5.21%
Symantec Corporation	197	3.65%
Amazon	190	3.53%
CloudFlare, Inc.	157	2.91%
Google Trust Services	148	2.75%
Others	643	11.93%

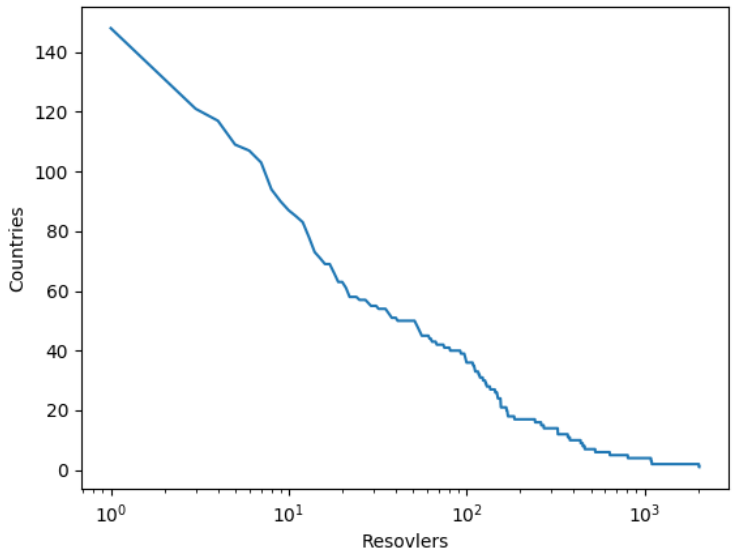


对 cn 顶级域名下证书服务商进行考察，情况与上表类似。

开放解析器

对 ZMap 检测到的 1.6 万个解析器进行数据分析：

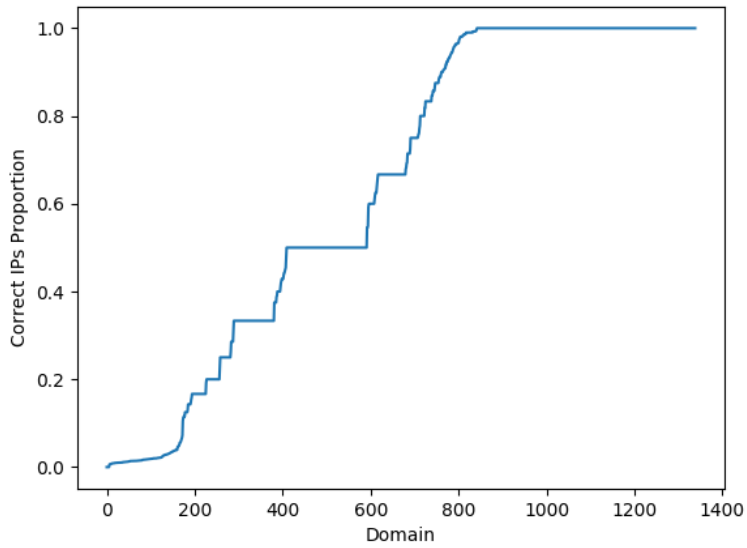
countryName	counts	DNSSEC	Percent
United States	2028	266	13.12%
Republic of Korea	2022	60	2.97%
Taiwan	1098	41	3.73%
Russia	1087	90	8.28%
Indonesia	806	27	3.35%
Japan	638	30	4.70%
United Kingdom	529	35	6.62%
Poland	462	30	6.49%
France	450	52	11.56%
Brazil	438	61	13.93%



解析正确性判断

- 1373 个域名在 16488 个解析器上进行查询，最终获得 926 万个 (domain, res, ip) 三元组，16 万 (domain, ip) 二元组
- 右图为联合分析算法的 `IPTrust` 结果与 HTTPS 验证结果进行对照的图像
- 一些国家的解析器会使用某些特殊 IP 地址作为对很多不相干域名的查询的响应，如左边表格所示

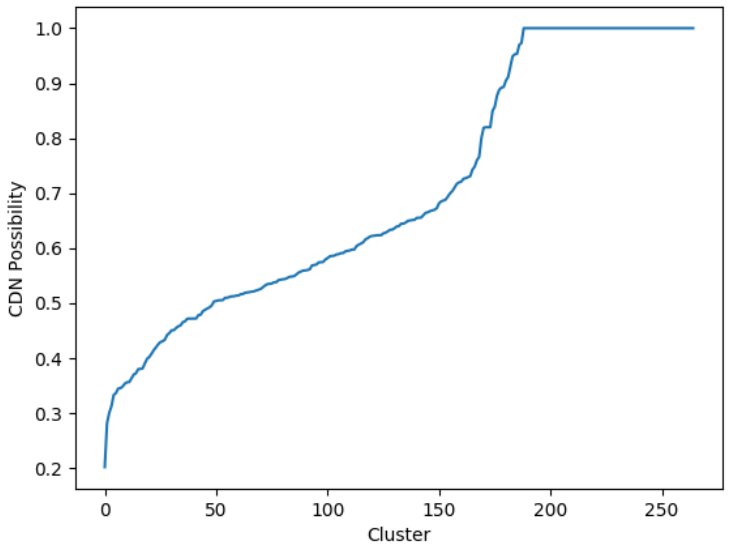
Country	IP	Counts	Domain Example
IRAN	10.10.34.34	131	bilibili.com
UAE	10.10.34.35	32	sex.com
FINLAND	0.0.0.0	27	exosrv.com
PORTUGAL	0.0.0.0	19	badoo.com
ROMANIA	0.0.0.0	11	cnzz.com
LATVIA	0.0.0.0	11	duba.com
SERBIA	79.101.14.184	7	www.goal.com
B&H	127.42.0.0	7	hatenablog.com
CYPRUS	127.42.0.0	6	iqoption.com
DENMARK	80.239.178.184	5	hm.com



CDN 识别判断

- 使用联合分析算法经过迭代得到 `DomainSimilarity` 矩阵
- 查找其中的连通分量，一个连通分量代表一个 CDN 的集群，最大的 CDN 集群的统计如下表所示
- 我们得到了数百个集群，通过验证对应组织名称的相似性对 CDN 识别算法进行验证，结果如右图

CDN	Size	Domain
Google	119	google.com
Fastly	73	airbnb.com
Amazon	40	time.com
Akamai	20	ebay.com
CloudFlare	10	4chan.org
AliCloud	7	tmall.com
Verizon	6	bing.com
Incapsula	6	prothomalo.com



参考文献

Marc Kührer, Thomas Hupperich, Jonas Bushart, Christian Rossow, and Thorsten Holz.

Going Wild : Large-Scale Classification of Open DNS Resolvers.

In *ACM Internet Measurement Conference (IMC), 2015*, pages 355–368, Tokyo, Japan, October 2015.

Will Scott, Sidney Berg, Thomas Anderson, Tadayoshi Kohno, and Arvind Krishnamurthy.

Joint Analysis of CDNs and Network-Level Interference using Satellite.

In *Proceedings of the 2016 USENIX Annual Technical Conference*, Denver, CO, USA, June 2016.

Paul Pearce, Ben Jones, Frank Li, Roya Ensafi, Nick Feamster, Nick Weaver, and Vern Paxson.

Global Measurement of DNS Manipulation.

In *Proceedings of the 26th USENIX Security Symposium*, Vancouver, BC, Canada, August 2017.


Thank you!