# 概述

论文主要做了两个贡献：

1、提出了三种方法来发现物联网设备的位置：流量中的服务器 IP 地址、DNS 查询中的服务器名称，这两个方法都利用了设备制造商运行的服务器的知识，了解物联网设备与之通信的服务器集；通过主动扫描获得的 TLS 证书，获得证书中的制造商信息；

2、将上述方法用于真实世界流量：基于IP地址的方法应用于采集的大学和IXP的流量，DNS的方法应用于8个根域名服务器的DNS流量，以研究自治系统的物联网部署。他们的发现包括：发现23种物联网设备的AS进入率增长了约3.5倍，使用这些设备类型检测到的AS的设备类型密度略有增加（2018年，在80%的AS中发现最多2种已知设备类型）。通过将基于DNS的方法应用于2013年至2017年某住宅小区的DNS流量，确认了家庭层面的IoT设备的部署的大幅增长。将基于证书的方法应用于公共TLS证书数据集，从199个国家找到254K 个IP摄像头和网络视频录像机（NVR）。

此外，论文公开了他们买的包含10个IoT设备的流量数据集，10个IoT设备列表如下：

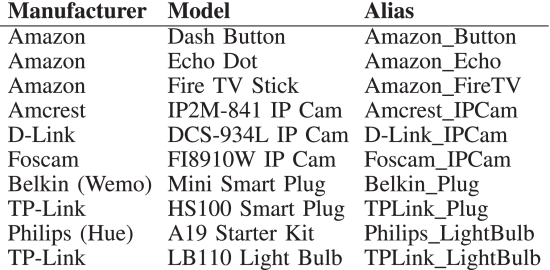


图1.该论文的IoT设备集

# 与先前工作的特色：

之前的工作大多需要从LAN口采集流量，并且不能推广到互联网范围的检测，本文的方法致力于观察NAT后面的设备以及公共IP地址上的设备；将方法应用于了大学校园、IXP、DNS根域名服务器等真实互联网环境，有助于全面地了解物联网设备在互联网上的部署情况。

# 发现IoT设备的方法

### A. 基于 IP 和 DNS 的检测方法

这两种方法都是从被动采集的流量中检测一般物联网设备。采集位置：对基于IP的方法：从互联网中任何有利的位置采集IP流量；对DNS方法：从存根服务器和递归服务器之间采集DNS查询流量。这两种方法利用了大多数物联网设备定期与设备特定服务器交换流量的观察结果，方法的核心思想都是1、先识别每个设备与之通信的服务器名称；2、通过 IP 地址或者 DNS请求中的域名与1、中已生成的服务器名称进行匹配。

知道服务器厂家，就可以通过观察与他们交换的流量来识别物联网设备的厂家，又由于服务器通常对于每类物联网设备都是唯一的，因此还可以识别设备的类型。该方法仅考虑物联网设备与谁交换流量，而没有时间或速率等特征，因为当流量混合时（例如NAT后面有多个设备）这些特征会被掩盖。

#### 1） 识别设备服务器名称：

通过购买的物联网设备样本（图1）并记录它们与谁交谈来生成候选服务器名称列表。对于我们购买的每个物联网设备，都会启动它并记录它发送的流量，从中提取服务器候选者的域名。我们在递归 DNS 服务器（recursive DNS resolver）的入口端捕获 DNS 查询，以减轻 DNS 缓存的影响。然后筛除第三方服务器（域名串匹配）和人为操作请求的（通过服务器是否响应重量级的Web请求）服务器，对于具有共享服务器的不同设备类型，没有进一步区分，而是保守地报告至少检测到其中一种设备类型，这称之为检测合并（detection merge），使用检测合并可以正确检测来自同一制造商的多个设备。

#### 2）基于IP的检测方法：

对于每种设备类型，我们跟踪设备类型到服务器名称的映射：一个设备类型与之对话的设备服务器名称列表。然后，我们定义服务器名称的阈值数量；将给定源ip（设备ip）的服务器名称阈值数的流量表示为某类型物联网设备的存在。对于这种方式，我们需要看到至少2/3的设备服务器名称（用设备服务器的ip地址标识），才能相信在给定的源IP地址上存在一种类型的物联网设备。

#### 3）基于DNS的检测方法：

通过在物联网设备和设备服务器之间进行实际的数据包交换之前识别DNS查询来检测物联网设备。探测方法上与2）类似，对于每种类型的 IoT 设备，跟踪设备类型与之通信的设备服务器名称列表，将来自给定 IP 地址的超出阈值的设备服务器名称数量的 DNS 查询的存在解释为该 IoT 设备类型存在。

该方法相比于2）直接根据ip地址查询的好处在于服务器ip地址可能改变，但域名不会变，且能自动添加后来出现的新域名，局限性在于此方法需要采集终端和递归 DNS 服务器之间的 DNS 查询。

### B. 基于证书的物联网检测方法

通过主动扫描TLS证书并识别目标物联网设备的TLS证书来检测使用HTTPS的物联网设备。值得一提的是Shodan等搜索引擎已经将TLS证书映射到物联网设备，方法是匹配证书文本内容，相比之下，我们的方法使用多种技术来提高证书匹配的准确性，并且还确认匹配的证书来自物联网设备中运行的HTTPS服务器。

首先确定候选证书：包含目标设备制造商名称和（可选）产品信息的TLS证书。候选证书很可能来自与目标设备相关的HTTPS服务器，例如制造商运行的HTTPS服务器和直接在其中运行的HTTPS服务器。然后，我们识别物联网证书：来自直接在目标设备中运行的HTTPS服务器的候选证书。

#### 1） 识别候选人证书：

根据与每个设备相关联的一组文本字符串（称为匹配密钥）测试每个TLS证书，从而确定每个目标设备的候选证书。

匹配密钥：如果目标设备的制造商不生产任何其他类型的互联网支持产品（根据制造商网站上的产品信息），则其匹配密钥仅为其制造商的名称（称为制造商密钥）。否则，其匹配密钥将是制造商密钥加上其产品类型（如“IP摄像头”）。

匹配方式：改进了字符串匹配的算法，考虑了一些现实因素定义了Good-Matches：密钥K是证书的某个域S的子串，但S中与K的匹配相邻的字符既不是字母，也不是数字。

匹配成功判断：对TLS证书C中的四个主题字段（组织CO、组织单元CUU、通用名称CCNN和ValueTaltNord CDN，如果存在），如果P的制造商密钥（KPM）Good-Matches了CO并且任何非制造商密钥与这四个主题字段中的任何一个相匹配，则认为证书C是设备P的候选证书。

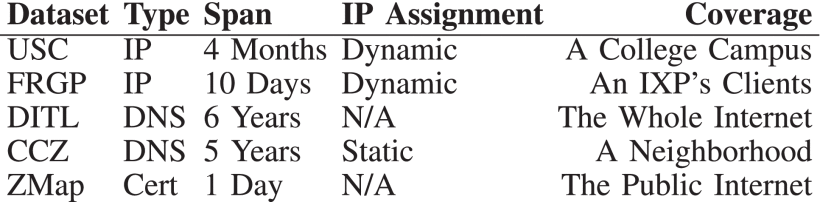
1）确定了每个设备的候选证书。

#### 2） 识别物联网证书：

识别物联网特定的证书，因为它们通常不是由证书颁发机构（CA）签署的。我们之所以识别它们，是因为它们是自签名的，并且缺乏有效的域名。IoT设备上的许多HTTPS服务器使用自签名证书，而不是CA签名证书，以避免CA签名的成本和复杂性。如果C的发行者组织CIO是我们检查的4个主题字段（Co，Coun，Cccn和CDN）的副本，或者由P制造商密钥（KPM）很好地匹配，我们考虑候选证书C（对于设备P）自签名。

# 结果：IoT Devices in the Wild

将上述IoT设备检测方法作用于下述真实环境下采集的公开数据集。



### 基于 IP 的物联网检测结果

识别设备服务器名称，这一步使用到的数据集是：他们自己购买的10台物联网设备（<https://ant.isi.edu/datasets/iot/> ）和来自新南威尔士大学提供的数据的21台物联网设备（<https://iotanalytics.unsw.edu.au/> ），他们手动检查 171 +48个候选服务器名称，并确认其中大多数的设备分类是否正确，最终留下它们总共有 99 个不同的设备服务器名称，用于区分23种设备类型。

然后将检测方法分别应用于大学环境和IXP流量。

对大学环境：

他们将方法作用于自己在“USC访客WiFi采集的4个月的流量”，得到一个USC在这四个月的IoT设备情况，比如发现 IoT IP 经常被检测到多种设备类型，这表明使用了网络地址转换 （NAT） 设备等。

对IXP上：

使用的公开数据集：FRGP Continuous Flow Dataset, May 2015, [online] Available: <http://www.isi.edu/ant/lander>. 对来自IXP的部分流量进行基于IP的检测，从111个IP中发现了122次10到11种设备类型的触发检测

### B. 基于 DNS 的物联网检测结果

将基于 DNS 的检测应用于两个实际的 DNS 数据集。

#### 1） 全球 AS 级物联网部署：

数据集：将检测应用于 2013 年至 2018 年的Day-in-the-Life of the Internet （DITL） 数据集，来显示在此期间 AS 级 IoT 部署的增长，每个 DITL 数据集都是一个根域名服务器在 2 天内接收的 DNS 查询。

#### 2）住宅区的物联网部署：

数据集：M. Allman, Case Connection Zone DNS Transactions, Jan. 2018, [online] Available: <http://www.icir.org/mallman/data.html>. 该数据集记录了俄亥俄州克利夫兰约100栋住宅的DNS查询。

上述两个检测结果分别报告了在AS级别可见物联网设备和居民区设备的识别情况。

### C. 基于证书的物联网检测结果

数据集：ZMap 443 HTTPS SSL Full IPv4 Datasets, Jul. 2018, [online] Available: <https://censys.io/data/443-https-ssl_3-full_ipv4>. 此数据集由 ZMap TCP SYN 扫描公共 IPv4 地址空间中的端口 443 上找到的证书组成。

检测结果：检测到9家制造商，最多的IP摄像机是大华等等。

# 验证方法

使用他们所拥有的设备进行受控实验，在他们自己的环境下采集数据对每种方法进行验证。比如对第一种方法基于IP的检测，通过将 10 个 IoT 设备和 15 个非 IoT 设备放置在家用路由器后面的无线 LAN 中来设置实验，将静态 IP 分配给这 25 个设备。我们在无线 LAN 内部运行 tcpdump，以观察从 LAN 到 Internet 的所有流量，运行了5天的实验，以模拟现实世界物联网测量中的可能情况。

