第51卷第6期
2019年12月

南京航空航天大学学报
Journal of Nanjing University of Aeronautics & Astronautics

Vol. 51 No. 6
Dec. 2019

DOI:10.16356/j.1005-2615.2019.06.019

一种基于多源数据的网络实体推断方法

马 昺[1] 仲思超[1] 蔡 冰[1] 王占丰[2]

(1.国家计算机网络与信息安全管理中心江苏分中心,南京,210003;
2.东南大学计算机科学与工程学院,南京,211189)

摘要: 网络空间中的实体推断是网络空间测绘研究的重要内容之一,主要通过综合多源数据实现对网络空间中各类实体的分类与识别。本文首先提出了网络空间的实体分类模型,基于此模型提出了一种低开销的网络实体探测分类方法。首先对于探测发现的IP地址,采用别名解析技术将属于一个设备的多个IP映射为一个网络实体;然后采用决策树对网络实体分类进行粗粒度分类;最后,再基于贝叶斯网络进行详细分类。为验证分类效果,以江苏省某市为例进行了探测分析并与备案数据进行了对比,试验结果表明该方法可以有效地对网络空间中的各类实体进行判别,从而为网络空间地图构建、态势分析等应用提供技术支撑。

关键词: 网络空间测绘;网络实体;贝叶斯;标注

中图分类号: TP393 **文献标志码:** A **文章编号:** 1005-2615(2019)06-0870-09

A Network Entity Inference Method Based on Multi-source Data

MA Yang[1], *ZHONG Sichao*[1], *CAI Bing*[1], *WANG Zhanfeng*[2]

(1. Jiangsu Branch of National Computer Network and Information Security Management Center, Nanjing, 210003, China;
2. School of Computer Science and Engineering, Southeast University, Nanjing, 211189, China)

Abstract: Network entity inference is one of the important contents in cyberspace surveying and mapping. Network entity inference and calibration need to synthesize multi-source data, classify nodes in network space by synthetical judgment. Thus, the entity classification model in network space is proposed. Based on this model, a low-overhead network entity detection classification method is proposed. Firstly, for IP addresses detected by detection, the alias parsing technology is used to map multiple IP addresses belonging to a device into a network entity; Then, the decision tree is used to classify network entities in a coarse-grained manner; Finally, the Bayesian network is used to classify them in detail. Taking a city in Jiangsu Province as an example, the detection and analysis are carried out and compared with the recorded data. The experimental results show that the method can effectively classify various types of entities in the network space, thus providing support for network space map construction, situation analysis and other applications.

Key words: network space surveying and mapping; network entities; Bayesian; annotation

伴随着信息技术的不断进步,网络已影响到政治、经济、文化、科技等各个方面,成为人类的"第二类生存空间"[1]。网络空间作为世界大国竞争的新疆域,网络空间测绘、网络空间安全等技术的研究受到学术界和产业界的重视,美国、俄罗斯等发达国家在政府部门设立了专门负责网络安全的机构和网络部队。网络空间测绘是感知网络空间态势,进行网络空间攻击和防御的前提,得到了高度的重视。网络空间测绘需要对网络空间中的各类实体及其属性进行探测、融合分析和绘制,从而形成对网络空间的全面认识。在此过程中,需要对网络实体进行分类,并对目标网络中的实体进行识别与

收稿日期: 2018-11-26;**修订日期:** 2019-09-12

通信作者: 马昺,男,高级工程师,E-mail:mayang@jsca.gov.cn。

引用格式: 马昺,仲思超,蔡冰,等. 一种基于多源数据的网络实体推断方法[J]. 南京航空航天大学学报,2019,51(6):870-878. MA Yang, ZHONG Sichao, CAI Bing, et al. A Network Entity Inference Method Based on Multi-source Data [J]. Journal of Nanjing University of Aeronautics & Astronautics,2019,51(6):870-878.

推断。

网络实体是指在网络空间中提供一定服务或具备一定功能的软件或者硬件。为了建立完善的网络空间实体地图，首先需要建立完善的网络空间分类图谱，然后通过网络测量的方法收集目标区域内各种网络的指纹，再进行网络实体分类。同时，伴随着云计算特别是虚拟化、NAT 等技术的大规模应用，单一的网络设备指纹探测工具，如 Nmap，Zmap，Masscan 等，无法实现网络实体的准确分类。因此，需要建立网络实体分类图谱，研究网络实体分类方法，并验证该方法在大规模真实网络中的有效性。

本文首先参照国内外相关研究对网络实体进行了分类，将其分为交换设备和端设备两大类，然后进一步细化形成了网络实体资源图谱。在此基础上，提出了一种低开销的网络实体探测分类(Low-overhead network entity detection and classification, LNEDC)方法，其基本思想如下：对于探测发现的 IP 地址，采用别名解析技术将属于一个设备的多个 IP 映射为一个网络实体；然后采用决策树的网络实体分类方法对网络实体进行粗粒度的分类；最后，再基于贝叶斯网络进行详细分类。在分类树构建过程中充分考虑网络测量开销的影响，从而使用较小的网络开销实现目标网络的快速探测与准确分类。为了验证算法的有效性，采用上述方法对江苏省某市的 IP 进行全面的探测与分类，并与备案库中的信息进行对比，从而验证算法的有效性。该算法的创新性在于能够对各类网络实体进行统一分类，而且具有较小的网络开销，从而为网络空间地图构建、态势分析等应用提供技术支撑。

1　相关工作

网络空间测绘是网络空间安全技术的重要组成部分，是进行网络空间安全攻防、网络安全态势评估的基础。美国是最早研究网络空间资源测绘的国家，经过近 20 年的建设，形成了较为完整的网络空间探测基础设施和体系，如美国国防局的 X 计划[2]，美国国土局的 SHINE 计划[3]，美国国安局的藏宝图计划[4]等。在国内，知道创宇公司的 Zoomeye、白帽汇的 FOFA 等建立了全球的路由设备、工业联网设备、物联网设备以及摄像头等基础设施的分类数据库。

在这些研究中主要分为两类，一类研究网络实体分类方法，一类是研究网络实体的识别。文献[1]中对网络空间测绘的概念、技术进行了分析，将网络空间中的实体分为实体资源和虚拟资源两大类，并对每一类进行了细分。文献[5]也采用了类似的分类方法，然而两者都没有对于网络实体分类进行系统的研究。Kohno 等[6]提出了采用设备的时钟偏差来进行硬件设备指纹识别，从而为推断设备类型提供依据。Lanze 等[7]提出采用时钟偏移来作为无线 AP 的标识，从而避免人为修改 SSID 和 MAC 地址导致的误判。采用相近研究思路的还有文献[8-10]。实际上，网络空间实体种类很多，还包括路由器、服务器、防火墙、运营商级 NAT(Carrier-grade NAT, CGN)等不同的实体，考虑到服务器的功能，还可以将其分为 DNS，Web，E-mail，流媒体服务器等。

在设备指纹分类识别方面，主要包括 3 类：基于 TCP/IP 协议栈指纹的 Nmap、基于设备指纹识别的 Shodan 以及设备指纹结合 Web 应用指纹的 Zoomeye。Nmap 主要使用 TCP/IP 协议栈指纹来准确地判断目标主机的操作系统类型；Shodan 主要采集设备开放端口的 banner(俗称“旗帜”)信息，通过 Banner 信息来进行分析，对设备信息进行标记；Zoomeye 结合了 Nmap 扫描工具，加强了对 Web 服务的识别能力，Zoomeye 团队通过工控协议的常用端口以及工控设备信息整理出了一个工控专题。Feng 等[11]提出一种可以通过自动学习来实现物联网(Internet of things, IoT)设备分类的方法，然而该方法仅适用于具有 HTTP 服务页面的物联网设备。

从上述分析可以发现，目前多数的网络实体探测或者分类研究主要具有如下特点和不足：(1)多数的工具和算法是针对一种设备或几种设备而设计的，缺少对于整个网络空间各种设备的统一分类方法；(2)目前的分类工具没有考虑多种网络实体分类时探测的顺序以降低探测开销。本文提出的 LNEDC 方法在建立网络空间实体统一分类的基础上，以降低网络测量开销为指导构建网络实体分类决策树并进行由粗到细的分类与识别，从而为构建网络空间态势地图提供依据。

2　网络实体分类

网络空间实体分类首先需要建立分类目录，然后采集网络实体指纹，并设计网络实体分类器，实现网络实体的分类与识别。因此，网络实体分类图谱的建立则是研究的基础，本节主要围绕网络空间探测的需要建立网络空间实体分类图谱。

网络实体具有不同的分类标准，文献[1，5]将网络空间中的所有实体分为两大类，即实体资源

和虚拟资源。实体资源分为交换设备和接入设备,虚拟资源分为内容、虚拟人和虚拟服务。实体资源是各种虚拟资源的载体,一个网络实体资源可能承载着多种虚拟服务,同时一个网络虚拟服务可能分布在多个物理实体上。为此,综合两种分类方法,结合网络空间实体探测与分类的需要,本文构建了如图1所示的分类方法图谱,将所有网络设备分为可以按照在网络中的位置和作用进行分类,如交换节点和端节点。其中,交换节点是指网络中的各种中间节点,为其他节点提供流量交换;端节点是指那些不为其他节点提供流量交换的节点。

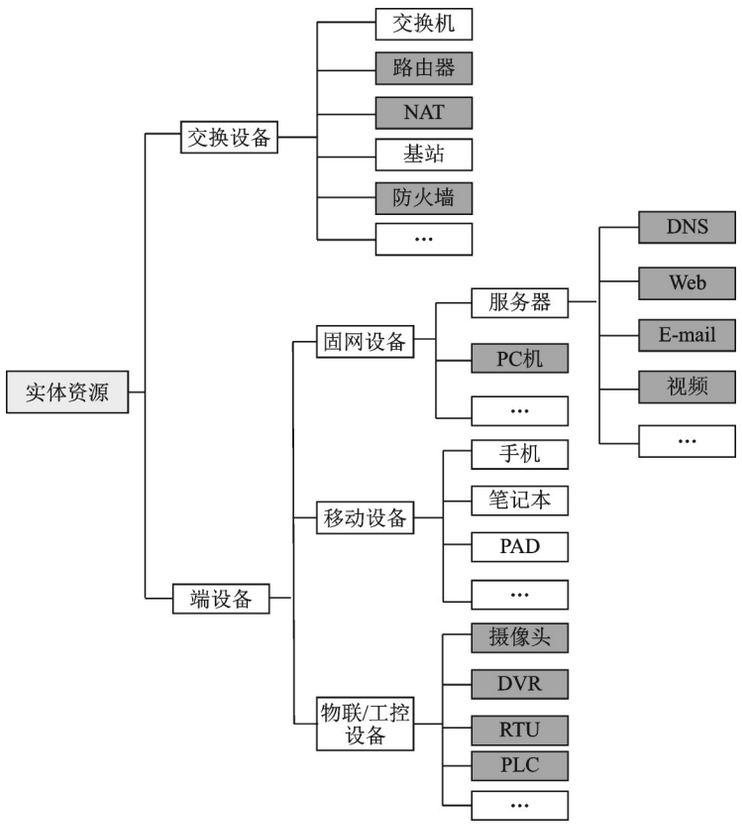


图1 网络实体分类图谱

Fig.1 Network entity classification map

图1中的所有网络实体并非都可以远程测量和分类,如交换机、基站等二层网络设备,而只有那些具有公网IP地址的设备才能被探测。因此,本文所研究的网络实体仅是那些可以通过网络测量获得的设备,主要包括路由器、防火墙、服务器、PC机、物联网设备、工控设备等,而基站、Wifi、交换机不在本文分析范围之内。此外,由于国内IP地址严重缺乏,互联网中采用了大量的NAT以及CGN技术,使得很多设备无法测量,从而影响了网络空间测绘的适用范围。为此,结合网络空间探测与分类的需要,进一步改进了网络空间分类图谱,构建了如图2所示的新型网络实体分类图谱。在此分类中,所有网络实体依然分为交换节点与端节点,与前面不同的是所有节点都是可以通过远程探测可以感知的。另外需要说明的是,随着当前网络功能虚拟化技术的发展,许多网络功能(如防火墙、路由器等)都不再必须在一个物理实体资源上实现,而是由虚拟机来完成,甚至一个物理实体资源还可以同时被虚拟化为多个不同的网络功能逻辑实体,在这种情况下,每个承载了网络功能的虚拟机都是一个网络实体,且该虚拟机根据其网络功能服务类型,归类到相应的属性类别中。

3 网络实体推断

在构建了完整的网络空间实体分类图谱后,就可以进行网络实体的探测与分类。这一过程又可以分为3个阶段,分别为网络实体发现、设备指纹的收集和网络实体识别与推断,下面分别对其进行详细说明。

3.1 网络实体发现

网络实体发现是从目标网络中筛选出活跃的IP地址作为推断的目标集合,并将属于同一个设备的多个IP映射为一个网络实体。实际上,网络实体的发现又可以分为3步:(1)目标地址集的提取,即从整个IP地址集中,按照区域、运营商等条件提出需要分析IP地址集合IP_{target};(2)多协议探测

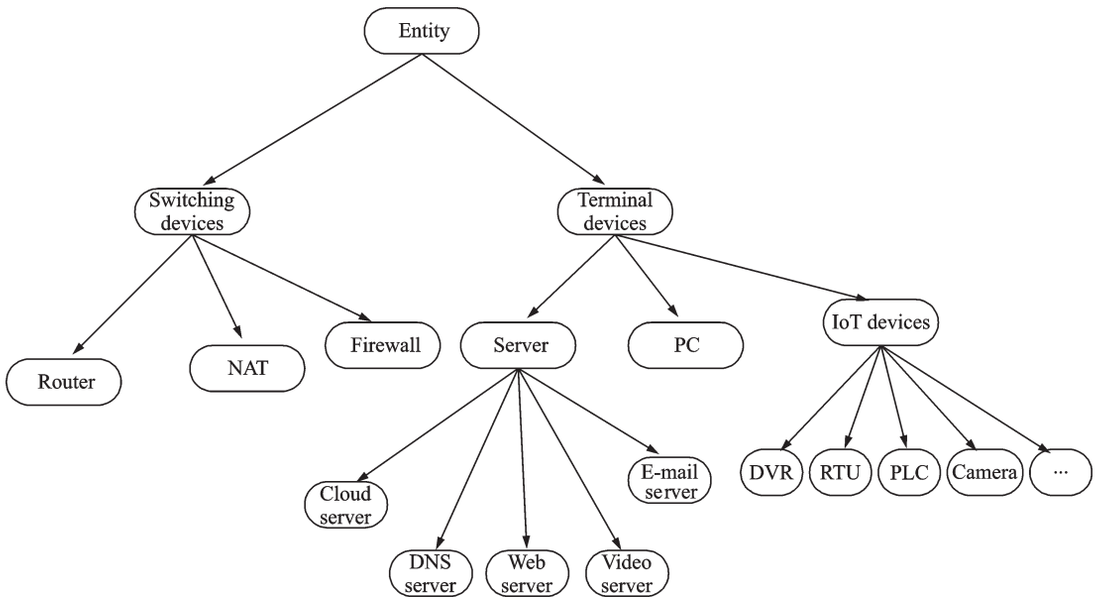


图 2　可探测网络实体分类图谱

Fig.2　Detectable network entity classification map

阶段，以往活跃网络实体，主要通过 Ping 的测量方式来完成，但是许多网络管理人员会禁止对于这种测量的响应，因此需要增加一些服务端口或者协议的探测[12]，为此选择与物联网设备、工控设备、典型服务相关的端口，采用 Nmap 进行低速探测，详细的端口列表和探测设置见 4.1 节；(3) 别名解析阶段，该阶段主要对多个活跃的 IP 地址进行分析，判定其是否配置在同一个设备上，别名解析的技术有很多种，经典的如 iffinder，Ally[13]，APAR[14]，Kapar[15]等，特别是由于 IPv4 与 IPv6 协议的差异，Speedtrap[16]，TBT[17]，TreeNet[18]等，本文在试验分析中选择了 iffinder 进行别名 IP 的分析与映射。

经过网络实体发现后，就建立了网络实体集合，接下来需要收集各种指纹信息，设计网络实体分类器，完成网络实体推断(图 3)。

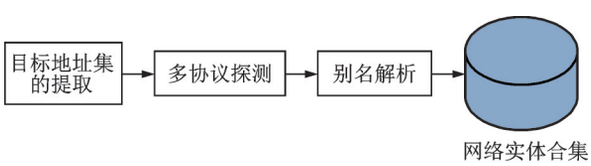


图 3　网络实体分析探测过程

Fig.3　Network entity analysis and detection process

3.2　设备指纹收集

设备指纹是网络设备识别的依据，不同的语境下设备指纹有很多种含义，如 Nmap 采用 TCP/IP 指纹辨识 OS；Shodan，Zoomeye，FOFA 等采用 HTTP 中的 banner 信息，有的还可以采用协议的交互信息。

对于一个网络设备 i 表示为 E_i，则该设备的属性集合 $F_i=\{f_1^i, f_2^i, \cdots, f_J^i\}$，其中 J 为属性的类别数量。实际上，在实际探测中，这个集合的特征数目要远远小于 J。对于设备判定的目标类别集合为 $G=\{g_k,\ 1<k<K\}$，设可以将设备进行分开的特征表示是 $F^k=\{f_1^k, f_2^k, \cdots, f_J^k\}$。因此，对于给定的分类集合 G，则其对应的探测集合特征为 F_G，为了减少探测开销需要对 F_G 进行约减。

设备的指纹或者属性信息包括网络位置、操作系统、开放端口、HTTP 协议 banner、协议信息，其基本含义如表 1 所示。在上述信息中，网络位置信息通过 traceroute 类的工具探测获得，操作性系统和开放端口采用 Nmap，Zmap 或 Masscan 来获取，HTTP 的 banner 信息获取也较为容易。物联网设备和工控设备的指纹获取较为困难，不仅数目众多而且往往采用私有协议。针对这类设备需要采用专有的协议，逐一进行探测和尝试，需要耗费大量的时间。目前，一些开源网站，如 Nmap、哈工大的灯塔实验室都有很多公开的工控探测脚本，如 S7，Modbus，SNMP，BACNet，EtherNet/IP，FINS，Fox，IEC，Moxa，dnp3，ProConOS，PcWorx，Cspv4

表 1　网络实体指纹特征说明

Tab.1　Description of fingerprint features of network entities

序号	特征名称	内容
1	网络位置	节点的活跃性
2	操作系统	被测目标设备的操作系统
3	开放端口	主机开放的服务
4	HTTP 协议 banner	主机的 banner 信息
5	协议信息	通过协议交互返回的信息

和Codesys等协议[19]。

设备指纹采集取决于网络实体分类的需要，采集的指纹信息越多带来的探测开销越大，不仅需要较大的存储空间，还会导致分类推断困难。为此，本文采取了一种分步递进测量的策略，首先对网络实体进行粗粒度分类，然后再进行细粒度分类。

3.3 网络实体识别与推断

在网络实体分类过程中，由于设备种类不同针对不同的设备需要采取特定的方法进行测量，因此必须采用分步测量与逐步推断的方式，下面对分步测量合理性进行证明。

命题 采用分步测量与推断将大大减小网络测量开销。

证明 首先假设网络中设备共分为M种N类，其中$M<<N$。假设对设备进行分类探测的报文开销，均为一致，对待测量IP地址集合$\mathrm{IP}_{\mathrm{target}}$，$|\mathrm{IP}_{\mathrm{target}}|\gg N$，则$C_{\mathrm{Overhead}}=\left(M+\sum_{i=1}^{\mathrm{IP}_{\mathrm{target}}}C_i\right)$，而全部测量的开销为$N$，因此$C_N\gg C_{\mathrm{Overhaed}}=\left(M+\sum_{i=1}^{\mathrm{IP}_{\mathrm{target}}}C_i\right)$。

在实际的网络当中，通过traceroute路径探测可以将网络实体判别为中间设备还是接入设备，其中中间设备均为路由器，通过Nmap扫描其系统指纹即可判断路由器归属厂商，而接入设备的种类相对更多，需要采用开放端口、操作系统指纹等信息来得到，因此可以根据网络实体的分类进行相应的补测即可。如果不采用分步测量的方法，而直接对所有设备进行所有信息的采集，无疑会带来更多的测量开销。同时，在所有的接入设备中工控和物联网设备十分丰富，而暴露在物联网中的设备又十分少，进行全谱探测的效率将会非常低，因此采用分步测量与逐步推断的方式将有效降低网络测量开销。

网络实体的识别主要依靠网络设备的指纹信息来实现，具体来说需要提取网络设备指纹中的一些关键信息，如开放的端口号、操作系统类型、协议关键词等，这些信息的分布概率与网络实体类别有较大的相关性，非常适合采用贝叶斯分类器来进行网络实体的推断。因此在网络实体分类过程中，将分类树与贝叶斯推断相结合，首先采用分类树进行粗粒度分类，然后采用贝叶斯网络进行细粒度分类。设一个网络实体i表示为E_i，其分类为G_i，则网络实体分类可以表示为：$G_i=\mathrm{Class}(E_i)$。由此，网络空间的实体识别已经转换为一个分类问题，需要设计一个多分类器。在分类器中，关键的是如何选择特征将不同网络实体进行区分，同时减少网络测量开销。设分类器的特征集合为$T=\{T_1,T_2,\cdots,T_n\}$，则对于特征T_i，待分类的IP地址集合C_{IP}，其测量开销为$M(C_{\mathrm{IP}},T_i)$，分类的准确性为$\mathrm{Positive}(C_{\mathrm{IP}},T_i)$。则网络实体推断的优化目标测量开销最小，同时判别最为准确。

$$\begin{cases}\min\left(\sum_{i=0}^{N}M(C_{\mathrm{IP}},T_i)\right)\\ \max\left(\sum_{i=0}^{N}\mathrm{Positive}(C_{\mathrm{IP}},T_i)\right)\end{cases}\tag{1}$$

3.3.1 粗粒度分类树的构建

在粗粒度分类中，将设备分为4类交换设备、服务器、PC机、物联网/工控设备。为了构建分类树，首先选择分类的特征信息，路由器与端设备的特征在于网络中的位置，表示为N_e；路由器与防火墙的区别在于操作系统指纹特征，表示为S；对于端设备进行分类主要依据其所提供的服务，表示为P，如表2所示。

表2 网络实体分类特征说明

Tab.2 Description of network entity classification features

序号	特征名称	符号	含义
1	边缘节点	N_e	是否为边缘节点，即是否出现在网络路径中
2	操作系统	S	操作系统类型与型号，判定是否为防火墙、服务器、物理网或工控设备等。
3	服务	P	主机开放的服务

按照上述特征依据网络实体图谱构建了网络实体的分类树，如图4所示。首先依据节点在网络中的位置判定是否路由器，凡是在路径库中不是位于最后一跳的IP均视为中间节点。考虑到测量数据的不完整性，依据备案数据建立路由器的操作系统数据库，根据采集的系统指纹信息，对端节点按照服务端口进行分类，从而建立起如图4所示的分类树。

3.3.2 基于贝叶斯的分类推断

在上述分类过程中，位于网络中间的路由器以及具有明显指纹特征的设备，判定是较为准确的。然而位于网络边缘的设备，由于系统指纹探测的非唯一性，往往无法判定其节点的类型。为此引入了贝叶斯判定网络，对设备类型进行分类。在分类中，依据系统的操作系统和开放服务作为判据。

设一个设备探测得到开放服务为$P_i=\{P_1,P_2,$

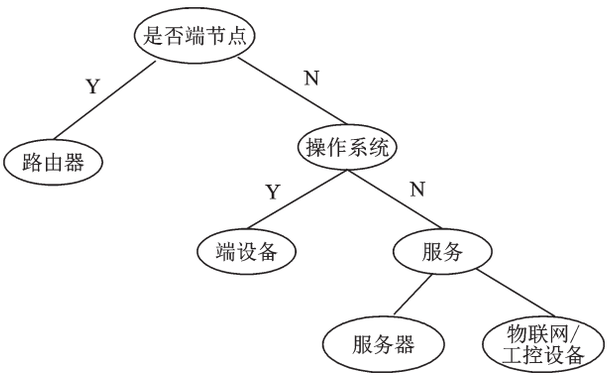


图4　网络实体分类树

Fig.4　Network entity classification tree

P_3, P_4, $\cdots$, $P_m\}$，则设备 E_i 为 G_i 的概率为

$$\tilde{c} = \mathrm{argmax}_{c_i \in C} P(c_i|P_i) \tag{2}$$

假设每种网络实体开放的服务或端口是独立的，根据贝叶斯理论，可以将式(2)表示为

$$P(M|c)=P(\{m_1, m_2, \cdots, m_n\}|c)= P(m_1|c)\cdot P(m_2|c)\cdot\cdots\cdot P(m_n|c) \tag{3}$$

概率 $P(m_i|c)$ 则是通过一维核密度来进行估计，该算法计算复杂度较小，同时具有较高精度。在实际的估算中，从IP备案中对所有报备IP进行了测量分析，作为先验概率，从而对其设备分类进行判定。

4　试验验证

为了验证上述算法的有效性，选择了江苏省某市的所有IP进行了测量与推断，总共的IP地址数目为5 367 770条，为了防止论文结果被用以网络攻击，本文不公开测试目标城市的名称。除了从CAIDA，Censys，iPlane采集的数据外，其余数据均采用OpenForm进行测量。OpenForm是国家互联网应急中心开发的开放式网络测量平台，集成了众多的开放方式网络系统资源，包括PlanetLab，Lookinglass等，同时加入了许多志愿节点，旨在提供分布式异构的网络测量资源池，系统支持二次开发。目前，该系统已集成节点超过1 000个，分布在50多个国家和地区，其中国内节点占到一半，是目前公开可用、分布最为广泛的测量平台。

4.1　测量数据说明

按照网络实体分类树的要求，需要采集的指纹主要包括网络路径、服务端口以及操作系统指纹。下面分别说明几个数据集的采集过程。

网络路径数据集：此数据主要对活跃的IP进行路径探测，从而获取网络路径信息，为减少测量开销，采集了CAIDA的数据集进行了过滤，同时考虑到测量位置不同得到的路径信息不同。在研究区域内选取了测量点，对于所有活跃的IP地址集进行了补充测量，将两者融合形成了一个统一的数据集。

服务数据集：为了发现更多的活跃网络实体，同时为实体分类提供指纹信息，选取了最常见的活跃端口进行了探测，主要涵盖Web，Email，DNS等13类通用服务，详见表3。

表3　服务探测及端口

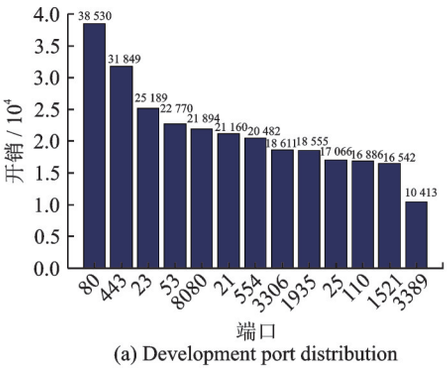
Tab.3　Service detection and port

序号	端口	服务	序号	端口	服务
1	80	HTTP	8	53	DNS
2	8080	HTTP	9	554	RTSP
3	443	HTTPs	10	1935	RTMP
4	21	FTP	11	1521	Oracle
5	23	Telnet	12	3306	Mysql
6	25	SMTP	13	3389	远程桌面
7	110	POP3			

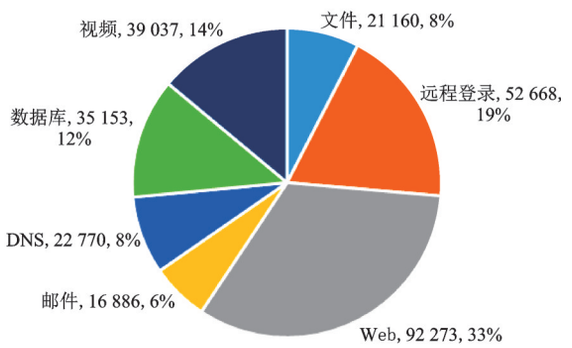
同样为了减少测量开销，采集了Censys数据集进行分析，同时在OpenForm平台中进行了补充测量。此次测量采用了分布于被测区域内部的3台计算机，主机分别位于电信、移动、联通3个运营商提供的机房。此次测量的城市共有备案IP地址5 367 770个，共发现活跃IP地址518 379个，约占报备IP总数的9.65%，探测发现服务端口279 947个，分布在77 047个独立的IP地址上，约为活跃IP的14.86%。图5(a)是测量中13个端口的分布情况，其中开放端口依次为80，443，23，53，8080端口等，主要为Web服务、文件服务以及DNS等网络基础服务。图5(b)按照服务的类型进行了归类统计，Web服务包括80，8080和443等，远程登录包括23，3389等，视频包括554，1935端口，由此可以发现在探测得到的IP中提供Web服务的最多，其次为远程登录服务、视频服务以及数据库服务，表明这些网络实体可能多数为服务器。

操作系统指纹：在OpenForm平台上使用10台主机采用Nmap对518 379个活跃的IP地址进行了主机指纹探测，共得到系统指纹30 877个，约占所有5.96%。在这些设备中，Linux服务器最多，其次为HP服务器、Huawei路由器、Windows主机以及Cisco路由等。从总体比例来看，单纯依靠系统指纹很难实现大规模设备的识别(图6)。

从上述分析可以发现，目标城市中备案IP为5 367 770个，响应IP约为其中9.65%，具有端口服务信息的约占其中的14.86%，具有操作系统指纹的约占5.96%。因此，可以进行全面识别分析的网络设备仅占其中很小一部分。



(a) Development port distribution



(b) Proportion of service types

图 5 服务探测数据统计

Fig.5 Service probe data statistics

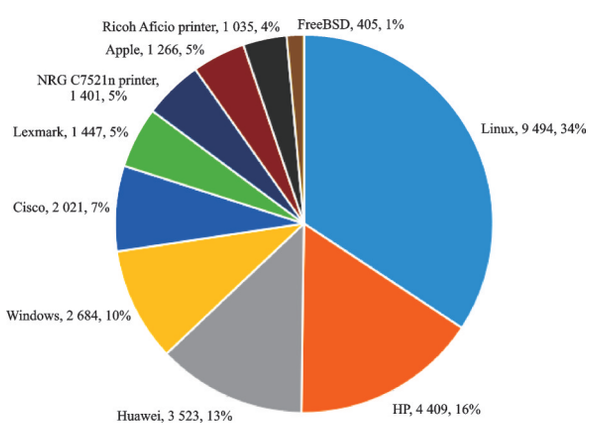


图 6 指纹探测数据统计

Fig.6 Fingerprint detection data statistics

4.2 网络实体探测

在网络中，IP 与网络识别并不是一一对应的，一个网络设备往往具有多个 IP，将 IP 与地理位置建立对应关系的技术称为别名解析技术。别名解析方法从原理上大体可以分基于测量的别名解析算法、基于推断的别名解析算法[20]以及综合的方法等 3 类。基于测量的别名解析算法向不同的地址发送探测分组，通过分析响应分组的相似性来判定两个 IP 是否属于同一个路由器。基于推断的别名解析算法通过构建 IP 拓扑图或者 IP 接口名称来推断 IP 地址是否属于同一个路由器。两种相比基于测量别名解析方法更为准确，但是对于不响应测量或者响应不完整的网络设备，后者也是一种补充。因此，综合的方法往往采用几种别名解析方法，从而做出更为全面和准确的判别[21]。在十几种别名解析算法中，采用了 iffinder 算法。其中，在分析中分为以下几个步骤：

(1) 选择江苏省某市 IP 地址共 5 367 770 个。进行活跃性测量后发现活跃 IP 地址 518 379 个。

(2) 对所有活跃 IP 进行 traceroute 测量，得到所有路径。

(3) 处于同一跳数上的并且地理位置在该市的 IP 地址存在别名解析的概率较高，对所有路径进行对比合并，结果 IP 共 9 148 个。

(4) 使用 iffinder 工具对结果集进行测量共得到 107 条存在别名解析的结果。

(5) 对结果进行合并得到 69 条最终结果。

经过上面的分析，得到了该市 IP 地址的别名数据，其中共发现具有 IP 的设备 69 台(图 7)，其中多数路由器具有一个 IP 别名地址，而只有少数的具有多个别名。经过别名解析将 IP 地址与设备进行了映射，其余没有别名的 IP 则单独视作一个设备，因此共发现 77 047 个设备。

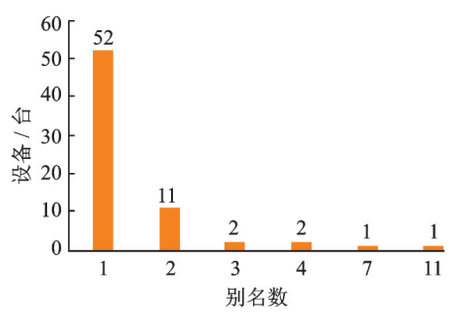


图 7 IP 别名个数

Fig.7 Number of IP aliases

4.3 网络实体分类结果

在对所有主机进行别名解析后，发现可以进行综合推断的网络实体 77 047 个。按照 3.3.1 节中网络实体分类方法分别按照网络设备进行了分类，发现其中 5 152 个为网络设备，其余的为端设备。对于剩余的 71 895 个网络实体，依据其操作系统以及服务端口号，进行了分类。

在所有的服务器中，仅提供一项服务的主机 IP 地址为 35 655 个。换言之，有 41 392 个 IP 地址提供两种以上的服务，从而为网络实体的推断带来了困难。考虑到目前许多系统的管理端往往采用 Web 作为前端，因此，当出现的 Web 服务和其他服务时，则判定为其业务系统。仅出现 Web 服务或者与数据库服务一起出现时，则判定为 Web 服务，这是由于 Web 服务往往需要数据库服务的支撑。

基于上述推断规则，采用了贝叶斯网络进行了训练，最终的分析结果如表4所示。

表4　网络实体推断结果

Tab.4　Inference results of network entities

序号	种类	数目	正确率/%
1	路由器	5 152	96
2	Web服务器	47 239	98
3	数据库	2 336	95
4	邮件服务器	2 595	86
5	DNS服务器	5 108	
6	视频服务器	7 032	60
7	文件服务器	7 585	82
	总数	77 047	

对分析结果结合备案信息库进行了验证，由于探测数据不够完整，因此仅对其正确率进行判别。从分析结果看，发现推测方法具有较高的正确率，但是召回率较小。DNS记录无法判定是由于目前的备案数据中没有完整的DNS记录，无法判定其正确性，但是从报文解析的完整性看，探测结果是可信的。经过分析出现误判的情况是某些服务器可能为云服务器，仅是开放了服务端口而没有提供相应的服务。

同时，为了验证本文所提出的网络实体推断方法的有效性，与Zoomeye和FOFA数据库中的结果进行了分析。由于Zoomeye和FOFA没有公开其具体采用的推断方法，本文根据测量城市的IP备案数据，从这两个数据库中抽取相关的识别结果，并从识别结果的覆盖率和正确率方面来进行比对，在排除无法进行正确性验证的DNS数据后，具体结果如表5所示。

表5　实体推断结果对比

Tab.5　Comparison of entity inference results

数据库	识别数目	正确率/%
OpenForm	71 939	92
Zoomeye	70 801	90
FOFA	68 930	87

从表5中的结果可以看出，无论是识别出的网络识别数目还是识别的准确率，本文所采用的方法都相对更高，这也说明LNEDC探测与分类方法对于准确推断网络实体的类型具有较好的实用效果。

5　结　　论

网络空间测绘是网络空间安全领域重要的研究方向之一，相对于传统的网络测量其内涵更加丰富，涉及的技术多而且复杂，然而由于其对网络空间安全态势分析、网络安全防御与侦查具有重要的作用，而备受关注。本文在现有研究成果基础上，结合主动探测的特点提出了可探测的网络空间实体分类图谱，基于此模型提出了一种低开销的网络实体探测分类方法LNEDC。以江苏省某个城市为例进行了探测分析并与备案数据进行了对比，试验结果表明该方法可以有效的对网络空间中的各类实体进行分类，但是存在召回率较低的问题，因此在今后需要研究如何提供主动探测的全面性，从而为构建更加全面的网络空间地图奠定基础。

参考文献：

[1] 郭莉，曹亚男，苏马婧，等. 网络空间资源测绘：概念与技术[J]. 信息安全学报，2018，3(4)：1-14.
GUO Li，CAO Yanan，SU Majing，et al. Cyberspace resources surveying and mapping：The concepts and technologies[J]. Journal of Cyber Security，2018，3(4)：1-14.

[2] NAKASHIMA E. With plan X，pentagon seeks to spread US military might to cyberspace[N]. Washington Post，2012-05-30.

[3] GRANT T. Leading issuesin cyber warfare and security：Cyber warfare and security[J]. On the Military Geography of Cyberspace，2015，2：119.

[4] RASHID O，MULLINS I，COULTON P，et al. Extending cyberspace：Location based games using cellular phones[J]. Computers in Entertainment，2006，4(1)：3-20.

[5] 赵帆，罗向阳，刘粉林. 网络空间测绘技术研究[J]. 网络与信息安全学报，2016，2(9)：1-11.
ZHAO Fan，LUO Xiangyang，LIU Fenlin. Research on cyberspace surveying and mapping technology[J]. Chinese Journal of Network and Information Security，2016，2(9)：1-11.

[6] KOHNO T，BROIDO A，CLAFFY K C. Remote physical device finger printing[J]. IEEE Transactions on Dependable and Secure Computing，2005，2(2)：93-108.

[7] LANZE F，PANCHENKO A，BRAATZ B，et al. Clock skew based remote device fingerprinting demystified[C]//Global Communications Conference. [S.l.]：IEEE，2012.

[8] LIBOR P，BARBORA F. On reliability of clock-skew-based remote computer identification[C]//International Conference on Security & Cryptography. [S.l.]：IEEE，2016.

[9] HUANG D J，YANG K T，NI C C，et al. Clock

skew based client device identification in cloud environments[C]//IEEE International Conference on Advanced Information Networking & Applications. [S.l.]: IEEE, 2012.

[10] KIKUCHI H, TOMINAGA Y, TANAKA Y. Remote host fingerprinting based on clock skew[C]// International Symposium on Communications & Information Technologies.[S.l.]: IEEE, 2008.

[11] FENG Xuan, LI Qiang, WANG Haining. Acquisitional rule-based engine for discovering internet-of-things devices [C]//27th USENIX Security Symposium (USENIX Security18).[S.l.]: [s.n.], 2018: 327-341.

[12] GASSER O, SCHEITLE Q, FOREMSKI P, et al. Clusters in the expanse: Understanding and unbiasing IPv6 Hitlists[J]. IMC, 2018,28: 364-378.

[13] SPRING N, MAHAJAN R, WETHERALL D. Measuring ISP topologies with rocketfuel[C]//ACM SIGCOMM.[S.l.]: ACM, 2002.

[14] GUNES M, SARAC K. Resolving IP aliases in building traceroute-based internet maps[J]. IEEE/ACM Transactions on Networking, 2009, 17(6): 1738-1751.

[15] KEYS K. Internet-scale IP alias resolution techniques [J]. ACM SIGCOMM Computer Communication Review, 2010, 40(1): 50-55.

[16] LUCKIE M, BEVERLY R, BRINKMEYER W, et al. Speedtrap: Internet-scale IPv6 alias resolution [C]//ACM Internet Measurement Conference (IMC).[S.l.]: ACM, 2013: 119-126.

[17] BEVERLY R, BRINKMEYER W, LUCKIE M, et al. IPv6 alias resolution via induced fragmentation[C]//Conference on Passive and Active Measurement.[S.l.]: [s.n.], 2013: 155-165.

[18] GRAILET J F, DONNET B. Towards a renewed alias resolution with space search reduction and IP fingerprinting[C]//Network Traffic Measurement and Analysis Conference.[S.l.]: IEEE, 2017:1-9.

[19] 灯塔实验室. 工业物联网安全态势分析报告.[EB/OL].(2019-05-12)[2019-08-09].http://plcscan.org/blog/.

[20] SPRING N, MAHAJAN R, WETHERALL D. Measuring ISP topologies with rocketfuel[C]//ACM SIGCOMM.[S.l.]: ACM, 2002.

[21] GRAILET J F, DONNET B. Towards a renewed alias resolution with space search reduction and IP fingerprintion[C]//Network Traffic Measurement and Analysis Conference.[S.l.]: IEEE, 2017: 1-9.

(编辑:夏道家)