MLAR：面向 IP 定位的大规模网络别名解析

袁福祥[1,2]，刘粉林[1,2]，刘翀[1,2]，刘琰[1,2]，罗向阳[1,2]
（1. 信息工程大学网络空间安全学院，河南 郑州 450001；
2. 数学工程与先进计算国家重点实验室，河南 郑州 450001）

摘　要：为准确高效地对接口 IP 进行别名解析，支撑 IP 定位，提出一种大规模网络别名解析算法（MLAR）。基于别名 IP 与非别名 IP 的时延、路径、Whois 等的统计差异，设计过滤规则，在解析前排除大量不可能存在别名关系的 IP，提高解析的效率；将别名解析转化为分类，构建时延相似度、路径相似度等四维新颖的特征，用于过滤后可能的别名 IP 和非别名 IP 的分类。基于 CAIDA 百万级样本的实验表明，相比 RadarGun、MIDAR、TreeNET，正确率提高 15.8%、4.8%、5.7%，耗时最多降低 77.8%、65.3%、55.2%；在应用于 IP 定位时，SLG、LENCR、PoPG 这 3 种典型定位方法的失败率降低 65.5%、64.1%、58.1%。

关键词：别名解析；IP 定位；网络拓扑；网络测量；机器学习
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.2096−109x.2020044

MLAR: large-scale network alias resolution for IP geolocation

YUAN Fuxiang[1,2], LIU Fenlin[1,2], LIU Chong[1,2], LIU Yan[1,2], LUO Xiangyang[1,2]
1. School of Cyberspace Security, Information Engineering University, Zhengzhou 450001, China
2. State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China

Abstract: In order to accurately and efficiently perform alias resolution on interface IP and support IP geolocation, a large-scale network alias resolution algorithm (MLAR) was proposed. Based on the statistical differences in delays, paths, Whois, etc. between alias IP and non-alias IP, before resolution, filtering rules were designed to exclude a large number of IPs that can not be aliases and improve efficiency of resolution, alias resolution was transformed into classification, and four novel features such as delay similarity, path similarity, etc. were constructed for the

收稿日期：2020−01−25；**修回日期：**2020−03−19
通信作者：袁福祥，rookieyfx@163.com
基金项目：国家自然科学基金（U1636219，U1736214，U1804263）；国家重点研发计划（2016YFB0801303，2016QY01W0105）；河南省科技创新杰出人才计划（184200510018）
Foundation Items: The National Natural Science Foundation of China (U1636219, U1736214, U1804263), The National Key R&D Program of China (2016YFB0801303, 2016QY01W0105), The Plan for Scientific Innovation Talent of Henan Province (184200510018)

论文引用格式：袁福祥, 刘粉林, 刘翀, 等. MLAR：面向 IP 定位的大规模网络别名解析[J]. 网络与信息安全学报, 2020, 6(4): 77-94.
YUAN F X, LIU F L, LIU C, et al. MLAR: large-scale network alias resolution for IP geolocation[J]. Chinese Journal of Network and Information Security, 2020, 6(4): 77-94.

classification of possible alias IP and non-alias IP after filtering. Experiments based on millions of samples from CAIDA show that compared with RadarGun, MIDAR, and TreeNET, the accuracy is improved by 15.8%, 4.8%, 5.7%, the time consumption can be reduced by up to 77.8%, 65.3%, and 55.2%, when the proposed algorithm is applied to IP geolocation, the failure rates of the three typical geolocation methods such as SLG, LENCR, and PoPG are reduced by about 65.5%, 64.1%, and 58.1%.

Key words: alias resolution, IP geolocation, network topology, network measurement, machine learning

1 引言

准确地刻画路由器级网络拓扑，对于分析网络的结构特性、感知网络的动态变化等十分重要[1-3]。现有的许多 IP 定位方法如 SLG[4]、LENCR[5]、PoPG[6]等往往依赖于路由器、网络地标（经纬度已知的稳定公网 IP）及待定位目标间的连接和时延关系，对目标 IP 实施定位。由于商业隐私保护等，路由器间的真实连接情况及对应的拓扑难以获取，研究者通常通过主动探测的方式进行推测。但路由器往往有多个接口，每个接口至少配置一个 IP[7]，这些 IP 互为别名关系，通过探测获取到的拓扑为 IP 接口级网络拓扑，而非实际的路由器级拓扑，因此无法满足基于路由器级拓扑的 IP 定位的需求。为了将 IP 接口级网络拓扑转化为路由器级网络拓扑，需进行别名解析，即分析哪些 IP 存在别名关系，判定哪些接口 IP 实际属于同一台路由器。开展对路由器别名 IP 进行准确、高效的发现及识别技术研究，对于获取真实的路由器级网络拓扑，进而利用拓扑中节点间的连接关系准确地定位目标 IP、追踪敏感用户、维护网络空间安全具有重要意义[8-15]。

现有典型的别名解析方法分为基于主动探测和基于被动分析两类。基于主动探测的方法主要通过对接口 IP 的探测获取响应报文，并基于响应报文首部源地址字段、标识字段以及可选字段等特点，进行别名解析；基于被动分析的方法则基于路由器主机名的命名规则、IP 地址指派惯例及网络构成，以及网络图结构等分析结果，进行别名解析。

典型的基于主动探测的别名解析方法如下。① 基于响应报文首部源地址字段的方法（如 Mercator[16]、Iffinder[17]等），利用对路由器接口 IP 进行 UDP 高端口探测时，响应报文首部的源地址字段可能出现另一接口 IP 这一特性，通过对比探测目的 IP 和响应报文中的源地址 IP，进行 IP 间别名关系判别。对该方法实际测试时发现，只有约 66%的目的 IP 地址响应 UDP 高端口探测，这其中只有 23%的地址返回原始目的 IP 以外的接口 IP。② 路由器多个接口 IP 通常共用唯一的计数器，每产生一个报文，计数器会在报文首部的 IP 标识字段（IP-ID，IP identification）设定相应的数值，若报文是连续的，该 IP-ID 值往往连续且线性增加。基于标识字段的方法则根据该特点，对可能存在别名关系的 IP 在较短的时间内相继发送多个请求报文，通过分析不同的响应报文中的 IP-ID 值，进行别名解析。例如，Ally[18]认为如果来自两个 IP 的响应报文中 IP-ID 值有序并且邻近，则该两个 IP 为别名；RadarGun[19]认为两个 IP 的多个响应报文中 IP-ID 序列较为相似，则该两个 IP 为别名；MIDAR[20]则认为当 IP-ID 序列的单调变化趋势相似时，两个 IP 为别名。但 RadarGun 的作者指出，在测试中，只有 31%的接口 IP 地址共享计数器；MIDAR 中指出，仅约 80.6%的接口 IP 会对探测返回可用于单调变化趋势判别的 IP-ID 序列。③ 基于可选字段的方法如 SideCar[21]、RIPAPT[22]、Pythia[23]等，则分别利用该字段可设置如记录路由、时间戳等报文控制信息，并依据记录结果中的接口 IP、时间戳等信息对 IP 进行别名解析。但 TreeNET[24]中指出，为了安全起见，大多数网络设备阻止数据包进行选项设置，一般会直接丢弃带有选项设置的报文。尤其自 2014 年 2 月以来，国际互联网工程任务组（IETF，The Internet Engineering Task Force）建议网络设备使用这种策略，致使这些方法几乎不再可用。

公开发表的基于被动分析的别名解析方法相对较少，代表性的有：基于路由器主机名命名规则的方法认为主机名相同或命名规则相似的 IP 为别名[25]，基于 IP 地址指派惯例及网络构成的方

法认为属于同一个/30 或/31 网段中的 IP 为别名关系[26]，基于图结构分析的方法则通过对接口 IP 间连接关系的分析进行别名解析[27]。然而，通过大量的测试发现，路由器主机名难以获取，命名规则不够规范，路由器存在大量未知接口（远超过 4 个），或无法得到接口 IP 间稳定的连接关系的情形十分普遍，导致在解析的准确性方面，仅基于主机名、/30 或/31 子网的 IP 分配，及图结构进行被动分析的别名解析方法不如基于主动探测的别名解析方法。

通过上述分析可知，在真实网络环境下，现有别名解析方法并不总能够获取到用于解析的相关数据，其准确性难以保证。研究者试图通过增加大量的探测或分析来解决该问题，但收效甚微，还引入大量的资源开销，同时大大降低了方法的效率。此外，在实际应用时，绝大部分现有别名解析方法在处理大量接口 IP 时，由于并不知道哪些 IP 之间存在别名关系，对任意一对 IP，这些方法往往需要对其进行别名关系判别，在别名解析前通过一系列特定的规则进行非别名 IP 过滤的方法非常少，个别典型方法如文献[28]的过滤效果仍然有待提高。这样，随着接口 IP 数量的增加，低效的别名解析难以适用于大规模网络。

上述问题的存在，使现有别名解析方法在实际应用时的准确率、效率一般，难以满足大规模网络的别名解析需求，从而影响了 IP 定位等实际应用的效果。例如，在使用 SLG、LENCR、PoPG 等基于路由器连接的目标 IP 定位时，由于无法准确高效地对大量的路由器接口 IP 进行别名解析，导致无法找到地标与目标 IP 间的共同路由器，不能根据地标位置估计目标 IP 的位置，从而造成基于共同路由器的定位方法失败。因此，有必要设计一种准确高效、适用于大规模网络的别名解析算法，以获取准确的路由器级网络拓扑，为目标 IP 定位等实际应用提供可靠的支撑。

针对上述问题，本文提出面向 IP 定位的大规模网络别名解析算法 MLAR（machine learning-based alias resolution）。MLAR 利用别名 IP 及非别名 IP 在直接时延、探测路径、Whois、主机信息等较容易获取的数据的统计差异，实现对大规模接口 IP 相对准确高效的别名解析，从而获取较为真实的路由器级网络拓扑，准确定位目标 IP。本文主要贡献及创新之处如下。

1) 提出了一种面向 IP 定位的大规模网络别名解析算法 MLAR。MLAR 可对大规模网络中的路由器接口 IP 进行准确、高效的别名解析，从而对大规模网络的路由器级网络拓扑进行准确的刻画，支撑基于网络拓扑的 IP 定位等应用。

2) 设计了提高别名解析效率的过滤规则。根据接口 IP 所属 ISP、探测路径及对应路由器主机信息的特性设计过滤规则，在进行别名解析之前，可依据规则排除不可能存在别名关系的 IP 对，从而减少别名解析的工作量，提高别名解析的效率。

3) 构建了用于对别名 IP 及非别名 IP 进行分类的四维特征向量。根据别名 IP 与非别名 IP 在直接时延、探测路径等较易于获取的数据方面的统计差异，将别名解析问题转化为机器学习中的分类问题，构建了分类特征向量，训练分类模型并用于对大规模网络的接口 IP 进行别名解析，提高别名解析的准确率。

2　别名 IP 与非别名 IP 的差异分析

本节对路由器接口 IP 的直接时延、探测路径、Whois 信息、IP 对应路由器主机信息等大量相关数据进行统计分析，并对别名 IP 与非别名 IP 在这些数据方面的统计差异进行详细介绍。

2.1　直接时延

从源 IP 向目的 IP 发送请求数据包，目的 IP 会对请求进行响应，通过该过程中目的 IP 对请求的响应时间，可得到源 IP 与目的 IP 之间的直接时延[29]。该时延与源 IP 和目的 IP 之间的距离有较大关系，在网络性能良好、拥塞不明显的情况下，地理距离越大，数据包在源 IP 与目的 IP 间传输消耗的时延越大[30-31]。同一源 IP 到处于相同地理位置的两个目的 IP 的时延往往相似，而到不同位置的目的 IP 的时延总会存在一定差异（除非目的 IP 大致分布于以源 IP 为圆心，源 IP 与目的 IP 距离为半径的圆上）。存在别名关系的 IP，配置在同一个路由器的不同接口上，其地理位置相同，因此两个别名 IP 相对于同一源 IP 的时延相似，而不存在别名关系的 IP 之间不具有这种

相似性。

本文获取了大量位于中国及美国的别名 IP 及非别名 IP 样本，从不同的探测源获取到每个 IP 的时延，并计算一对别名及非别名 IP 间的时延差值。分别取 10 000 对上述差值，图 1 为差值对比，其中“+”代表一对非别名 IP 间的时延差值，“○”则代表一对别名 IP 间的时延差值。图 1(a)中样本位于中国北京，探测源位于成都；图 1(b)中样本位于美国纽约，探测源位于亚特兰大。由图 1 可看出，同一探测源到一对别名 IP 的时延大多较为相似，差值较小，约小于 5 ms，而到一对非别名 IP 的时延相似程度较低，时延差值较大。由图 1 可得，从不同探测源获取别名 IP 间及非别名 IP 间的时延差值往往存在明显的差异，这种差异可用于区分别名 IP 与非别名 IP。

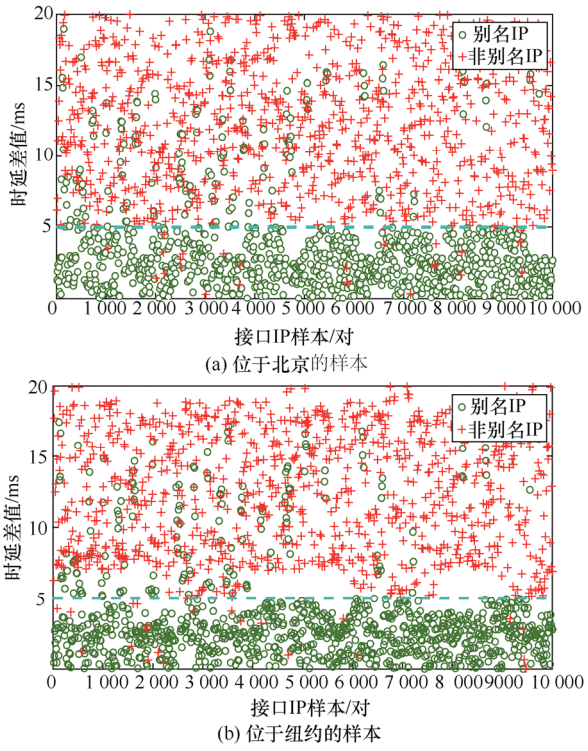

图 1　不同城市的别名 IP 与非别名 IP 时延差值对比
Figure 1　Comparison of delay difference between alias IP and non-alias IP in different cities

2.2　探测路径

路由器主要负责为转发的每个数据包寻找一条最佳的传输路径，从而将数据包高效地传送到下一跳。为了能够快速选择出最佳路径，路由器中保存了包含数据转发策略的路由表，供路由选择时使用[32]。通常，该路由表在相当一段时间内是不变的，即路由节点的下一跳是相对固定的。从源 IP 到目的 IP，通常会经过多个路由器，由于每个路由节点的下一跳在一段时间内相对固定，那么从源 IP 到目的 IP 的整条路径也是固定的。

别名 IP，被配置在同一路由器上，无论其地理位置，还是在拓扑中的逻辑位置都是相同的。根据上述路径的稳定性可知，从同一源 IP 到别名 IP 的探测路径应相同或极为相似，而到非别名 IP 的路径相似度应相对较低。利用 traceroute，获取从同一探测源到大量接口 IP 的探测路径。对这些路径进行分析发现，其相似程度可分为如下 A~D 这 4 种情况，图 2 为不同情况的示意。其中，由于路径的方向是由所经过的路由器决定的，因此，本当两条路径不同的路由 IP 数量小于或等于 2 时，路径的方向是相似的，当两条路径的跳数差异小于或等于 2 时，路径的长度是相似的。

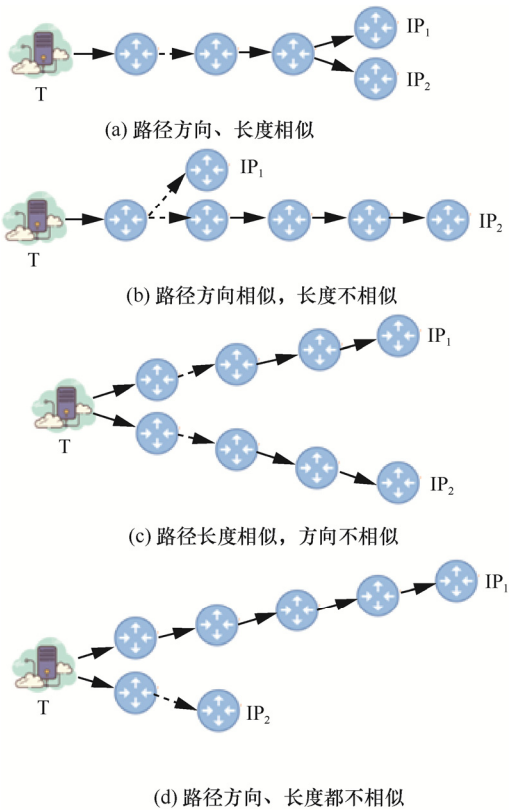

图 2　两个 IP 的探测路径相似性分析
Figure 2　Similarity analysis of probe paths between two IP

A：两个接口 IP 的探测路径方向、长度极为

相似。如图 2(a)所示，从探测源 T 到 IP_1，IP_2 的探测路径，跳数基本相同，且对应的每一跳基本是同一路由节点。

B：两个接口 IP 的探测路径方向相似，但长度不相似。如图 2(b)所示，从探测源 T 到 IP_1，IP_2 的探测路径，跳数差异较大，但初始的多跳路由节点基本相同。

C：两个接口 IP 的探测路径长度相似，但方向不相似。如图 2(c)所示，从探测源 T 到 IP_1，IP_2 的探测路径，跳数基本相同，但对应的每一跳几乎是不同一路由节点。

D：两个接口 IP 的探测路径方向、长度都不相似。如图 2(d)所示，从探测源 T 到 IP_1，IP_2 的探测路径，跳数存在一定差异，且对应的每一跳是不同一路由节点。

对于上述 4 种情况，A 中两个接口 IP 很大程度上存在别名关系，而 C、D 中两个 IP 一般不可能存在别名关系。对于 B，当探测路径跳数相差较少时，该两个 IP 可能存在别名关系，当路径跳数相差较大，如 3 跳及以上时，该两个 IP 几乎不可能存在别名关系，还有一种极端的情况是两个接口 IP 出现在同一条路径上，此时两个 IP 被配置在不同的路由器上，不可能存在别名关系。对 1×10^6 对别名 IP 及非别名 IP 的探测路径进行统计分析，对应不同路径相似程度的 IP 对所占比例如表 1 所示。

表 1 列出了别名 IP 及非别名 IP 的探测路径相似程度对应 A、B、C、D 的比例，以及每种情况下路径的方向及长度的不同相似程度的具体比例。由表 1 的统计结果可得，所有的别名 IP 对的路径相似程度都属于 A 或 B，但属于 A 的占 98.1%，而属于 B 的仅占 1.9%，且非别名 IP 对中，路径相似程度属于 A 的仅有 0.4%，而有 13.1%的属于 B，这说明两个 IP 探测路径相似程度属于 A 时，很大程度上可能互为别名，属于 B 时是否为别名存在一定的不确定性，而当属于 C 或 D 时，基本不会成为别名。这种别名 IP 及非别名 IP 在探测路径的方向、长度方面相似程度的差异，可用于过滤不可能存在别名关系的 IP，以及判别 IP 间是否存在别名关系。

2.3　Whois 信息

IP 的 Whois 信息，即 IP 的详细数据信息，主要包括 IP 所属单位描述、IP 的持有者及相关信息、信息最后修改时间等。存在别名关系的 IP，被配置在同一个路由器上，其 Whois 信息往往相同，而非别名 IP 的 Whois 信息差异较为明显。对 1×10^6 对别名 IP 及非别名 IP 的 Whois 信息进行统计，结果如表 2 所示，由表可以看出，至少有 98.4%的别名 IP 对的 Whois 信息基本一致，相同的信息条数≥15，而约 93.9%的非别名 IP 对的 Whois 信息仅有 4 项相似，如技术联系人、通信地址等。尽管非别名 IP 间也存在个别的信息项相同，但总体而言，别名 IP 对与非别名 IP 对在 Whois 信息相似程度方面的差异，可以为别名解析提供帮助。由于 IP 的 Whois 信息无须通过探测获取，仅通过查询 Whois 信息库即可得到，因此即使在待解析的路由器接口 IP 对探测无响应时，仍能够在一定程度上利用 Whois 信息进行 IP 间别名关系的判别。

表 1　路径相似程度统计

Table 1　Statistics of path similarity

IP 对	类别	比例	方向：路径中不同路由 IP 的数量			长度：探测路径的跳数差异		
			≤1	2	≥3	≤1	2	≥3
别名 IP 对	A	98.1%	95.2%	2.9%	0	96.4%	1.7%	0
	B	1.9%	1.4%	0.5%	0	0	0	1.9%
	C	0	0	0	0	0	0	0
	D	0	0	0	0	0	0	0
非别名 IP 对	A	0.4%	0.1%	0.3%	0	0.2%	0.2%	0
	B	13.1%	4.3%	8.8%	0	0	0	13.1%
	C	34.4%	0	0	34.4%	10.9%	23.5%	0
	D	52.1%	0	0	52.1%	0	0	52.1%

表 2 相同 Whois 信息项的条数统计
Table 2 Statistics of the number of identical Whois information items

相同 Whois 信息项的条数	别名 IP 比例	非别名 IP 比例
≤4	0.2%	93.9%
5~10	0.5%	3.7%
11~15	0.9%	1.6%
≥15	98.4%	0.8%

2.4 路由器主机信息

作为网络中重要的“枢纽”，路由器主要负责网络中数据包的转发。像计算机使用 Windows、Linux 等作为操作系统一样，在路由器上，也有软件在运行，可以等同地认为它们就是路由器的操作系统，这种系统主要负责完成路由表的生成和维护，如 FreeBSD、Juniper JUNOS、OpenBSD 等[33]。不同路由器的操作系统可能不同，存在别名关系的 IP 对应的路由器其操作系统一定相同。

为了提供多种服务，满足不同的网络需求，路由器会开放多个端口，不同的路由器开放的端口可能不一样，存在别名关系的 IP 对应的路由器，其开放的端口及对应端口的状态一定相同。此外，在相同时刻，存在别名关系的 IP 对应的路由器的运行状态（即在线或者离线）是一致的，而不存在别名关系的 IP 对应的路由器，可能由于断电或网络中断等导致其运行状态不一致。

同样地，分别对 1×10^6 对别名 IP 及非别名 IP 的主机信息进行统计后发现，约 89.8%的别名 IP 对应的路由器的操作系统信息一致，约 96.6%的别名 IP 对应的路由器的端口开放情况完全一致，所有的存在别名关系的一对 IP 对应的路由器的运行状态完全一致，而非别名 IP 对应的路由器上述信息一致的比例分别仅为 12.1%、6.9%、1.0%，差异较为明显。

IP 所属的互联网服务提供商（ISP，internet service provider）信息，也可用于判别 IP 间是否存在别名关系。配置在同一路由器上的 IP，往往属于同一 ISP（骨干网路由器除外，因为个别骨干路由器不同接口 IP 可能属于不同 ISP）。若某两个接口 IP 不属于同一 ISP，则该两个 IP 不存在别名关系。

通过上述统计分析可知，别名 IP 的直接时延、探测路径、Whois 信息、路由器主机信息等数据相似性较高，而非别名 IP 间的这种相似性往往较低，这些明显的差异可用于区分别名 IP 与非别名 IP。

3 MLAR 算法

基于第 2 节中给出的别名 IP 与非别名 IP 在直接时延、探测路径等方面存在的差异，提出了基于机器学习的别名解析算法 MLAR。MLAR 给出了一组非别名 IP 过滤规则，排除不存在别名关系的 IP，减少别名解析的工作量，提高别名解析的效率；MLAR 将别名判别问题转化为分类问题，将别名 IP 对作为正例样本，非别名 IP 对作为负例样本，构造了用于对别名 IP 对和非别名 IP 对进行分类的四维特征，对利用规则过滤后剩余的 IP 对进行别名解析。

3.1 MLAR 基本原理与主要步骤

MLAR 算法主要包括样本集合构造、相关数据获取、非别名 IP 过滤、分类特征表示等步骤，具体如下，其原理框架如图 3 所示。

输入 别名 IP 及非别名 IP 样本集 S，待解析路由器接口 IP 集 S'

输出 S'中接口 IP 的别名解析结果 R

Step 1 样本集合构造。从公开数据源或路由节点已知的网络获取特定目标区域内一定数量存在别名关系的接口 IP 对，构成集合 S_0；同时，获取一定数量的不存在别名关系的 IP 对，构成集合 S_1；总的样本集合 $S=S_0\cup S_1$。区域内待解析的所有路由器接口 IP 构成集合 S'。

Step 2 相关数据获取。分布式部署多个探测源，对集合 S 及 S'中的接口 IP 进行探测，获取从源 IP 到接口 IP 的时延和路径；通过查询相关 IP 信息库，获取每个接口 IP 所属 ISP 及 Whois 信息；通过探测源对接口 IP 对应的路由器主机进行监测，获取其操作系统版本、端口开放情况以及主机运行状态等信息。

Step 3 非别名 IP 过滤。对 S'中的任意接口 IP 进行两两组合，并利用 Step 2 获取的数据，对 IP 所属 ISP、探测路径及对应主机运行状态进行统计。根据设计好的过滤规则，排除不存在别名

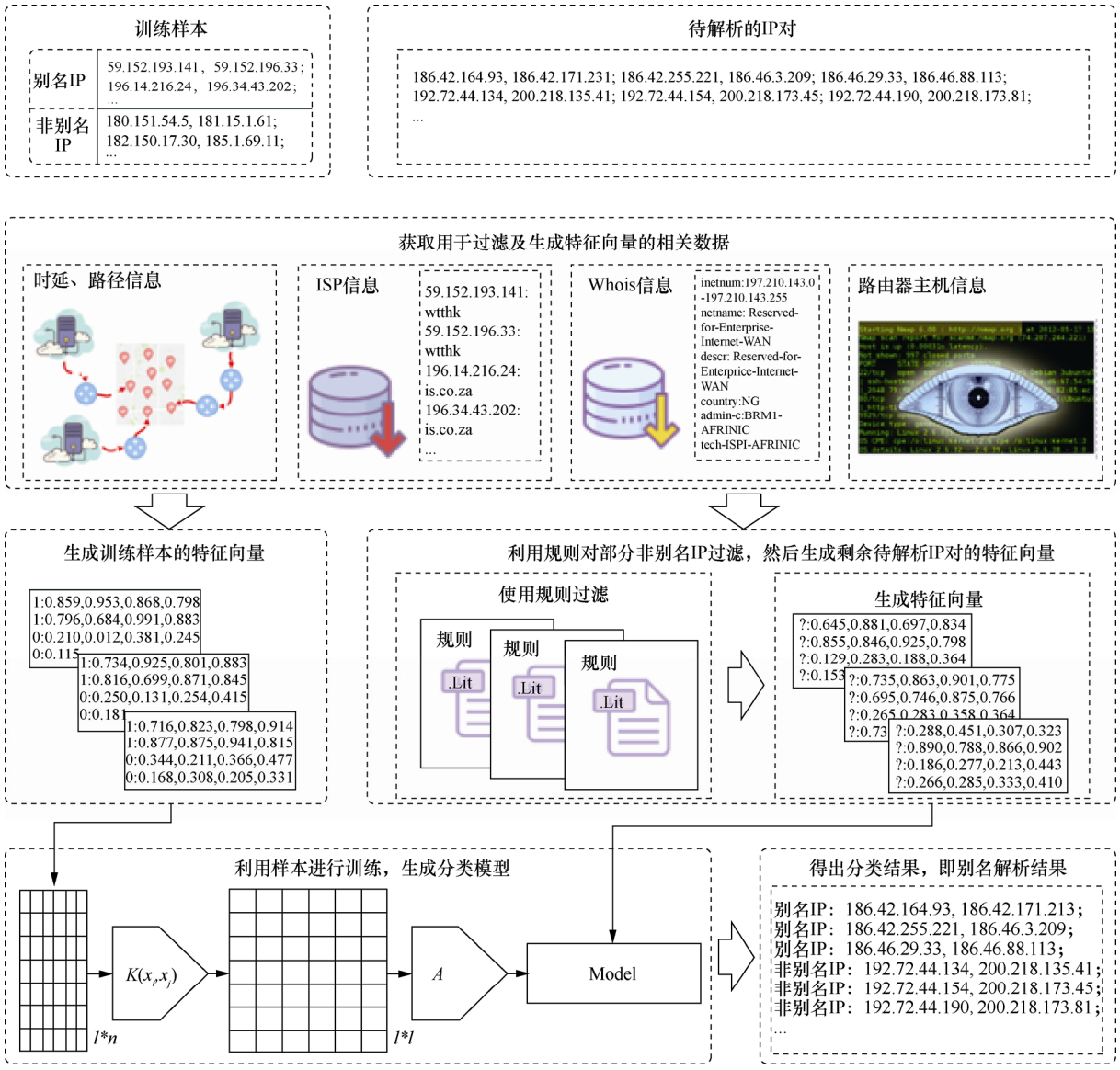

图 3　MLAR 的原理框架

Figure 3　The principle framework of MLAR

关系的接口 IP，剩余 IP 构成集合 S''。

Step 4　分类特征表示。$\forall(\mathrm{IP}_i, \mathrm{IP}_j)\in S\cup S''$，利用 IP_i, IP_j 的时延、探测路径、Whois、路由器主机等信息，根据设计好的分类特征生成方法，为样本$(\mathrm{IP}_i, \mathrm{IP}_j)$构造特征向量 $\boldsymbol{F}_{i,j}(F_1, F_2, F_3, F_4)$。获得 S 中所有 IP 对的特征向量，构造集合 F。同样，对于过滤后生成的集合 S''中的 IP 对，构造集合 F'。分类特征如表 3 所示。

Step 5　分类模型训练。不同的分类器特点不同，对相同样本的分类效率及效果会存在一定的差异（通常，线性分类器的效率相对较高。在线性分类器中，不同的分类器对数据缺失、噪声等因素的敏感程度不同）。为保证良好的分类效率，同时结合 Step 2 中所获取的相关数据的特点，如数据规模、数据缺失程度、噪声数据的比例等，选择合适的线性分类器。将特征向量集合 F 作为分类器的输入，对分类器进行训练，得到分类模型 Model。

Step 6　别名解析。对于集合 S''中待解析的 IP 对，将其特征向量集合 F'输入已训练好的模型 Model，得到分类结果 R，即任意一对 IP 的别名解析结果。

上述步骤中，非别名 IP 过滤及分类特征的表

示是算法最为关键的环节，3.2 节和 3.3 节将对这两部分分别进行具体阐述。

表 3 特征集合
Table 3 The feature set

特征序号	特征名称
F_1	时延相似度 S_d
F_2	路径相似度 S_p
F_3	Whois 信息相似度 S_w
F_4	主机信息相似度 S_h

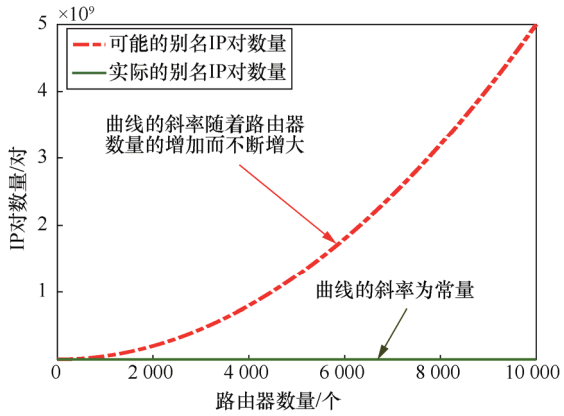

图 4 可能的别名数量与实际的别名数量对比
Figure 4 Comparison of the number of possible aliases with the actual number of aliases

3.2 非别名 IP 过滤

实际网络中的路由器，一般具有多个接口，每个接口会配置一个 IP。同一路由器上的多个接口 IP 之间互为别名关系，而不同路由器上的接口 IP 不存在这种关系。假设某个网络存在 P 个路由器，每个路由器有 Q 个接口 IP，在获取到该 PQ 个接口 IP 后，传统的别名解析方法直接对其中任意的两个 IP 进行组合，并对 IP 间是否存在别名关系进行判别，需要解析的 IP 对数量为 $C_{PQ}^2=\frac{PQ(PQ-1)}{2}$，但其中真正存在别名关系的 IP 对数量仅为 $PC_Q^2=\frac{PQ(Q-1)}{2}$，二者之比为 $\frac{PQ-1}{Q-1}$，P 越大，二者差异越明显。如图 4 所示，设每个路由器平均有 10 个接口，即上述 Q=10，红色曲线表示传统方法认为可能的别名对即需要解析的 IP 对数量，绿色曲线表示实际的别名对的数量，对比发现，随着路由器数量的增加，二者差异越来越大。

通过上述分析可知，若能够在别名解析之前，尽可能地过滤掉不可能存在别名关系的 IP 对，则可以减少别名解析工作量，显著提高别名解析的效率。基于第 2 节的统计分析，给出一组非别名 IP 过滤规则。对待判别 IP 对，通过如下规则进行过滤，排除不可能存在别名关系的 IP 对。设两个路由器接口 IP 分别为 IP_1、IP_2，IFalias(IP_1, IP_2) 表示判别 IP_1 与 IP_2 是否存在别名关系的布尔型函数，若 IFalias(IP_1, IP_2)=1，IP_1 与 IP_2 存在别名关系；若 IFalias(IP_1, IP_2)=0，IP_1 与 IP_2 不存在别名关系，则有如下规则。

1) 由于同一路由器的不同接口 IP 属于同一 ISP，因此不属于同一 ISP 的任意两个非骨干路由接口 IP 不存在别名关系，即设 ISP(IP)表示 IP 所属的 ISP，对于$\forall IP_1$、IP_2，若 $ISP(IP_1)\neq ISP(IP_2)$，则 IFalias(IP_1, IP_2)=0。

2) 探测路径中每一跳 IP 属于不同的路由器，因此出现在同一条探测路径中的两个接口 IP 不存在别名关系，即设 PATH 表示一条探测路径中的所有路由器 IP 构成的集合，对于$\forall IP_1$、IP_2，若 $(IP_1\in PATH)\wedge(IP_2\in PATH)$，则 IFalias($IP_1$, IP_2)=0。

3) 同一探测源到同一路由器不同的接口 IP 的路径方向相似，因此从同一探测源获取的任意两条路径，其相同跳的 IP 不同的情况出现次数大于或等于 3 时，两个接口 IP 不存在别名关系，即设 List(IP)表示 IP 的探测路径中所有中间路由器 IP 按从源 IP 向 IP_1 的顺序构成的集合，对于$\forall IP_1$、IP_2，设 C(List(IP_1), List(IP_2)) 表示 List(IP_1) 与 List(IP_2)的不同元素构成的集合，若有$|C(\text{List}(IP_1), \text{List}(IP_2))|\geqslant 3$，则 IFalias($IP_1$, IP_2)=0。

4) 同一探测源到同一路由器不同的接口 IP 的路径长度相似，从同一探测源获取的路径的跳数差异大于或等于 4 时，两个接口 IP 不存在别名关系，即设 Len_T(IP)表示从探测源 T 到 IP 的探测路径的跳数，对于$\forall IP_1$、IP_2，若$|\text{Len_T}(IP_1)-\text{Len_T}(IP_2)|\geqslant 4$，则 IFalias($IP_1$, IP_2)=0。

5) 在相同时刻，存在别名关系的 IP 对应同一台路由器，其运行状态是确定的，因此，对应主机运行状态不同的任意两个接口 IP 不存在别名关系，即设 Status_t(IP)表示 IP 对应的主机在特定 t 时刻的运行状态的布尔型函数，

Status_t(IP)=1，则主机在线，Status_t(IP)=0，则主机离线，若 Status_t(IP_1)≠Status_t(IP_2)，则 IFalias(IP_1, IP_2)=0。

对于待判别的 IP 集合及任意组合的一对 IP，使用上述规则，进行过滤。需要说明的是，以上规则是有先后顺序的。这是因为规则 1)中 IP 所属 ISP 可以通过查询现有的数据库获取；规则 2)、3)、4)为确保准确性，综合采用多个探测源并行探测，并根据探测结果进行判别，耗时较少；相对而言，规则 5)则需要对 IP 对应的主机监测一段时间，因此将其放在最后进行，且仅对通过规则 1)~4)过滤后的 IP 进行监测。由 2.2 节的分析及表 1 统计结果可知，通常探测路径的跳数差异大于或等于 3 时，两个 IP 基本不存在别名关系，但为了降低个别特殊 IP 对带来的误判，在规则 4)中，进一步将阈值增大为 4。在 MLAR 中，依据上述过滤规则，排除不存在别名关系的 IP 对后，对剩余的 IP 对，利用 3.3 节中设计的分类特征表示方法，生成特征向量，进行分类及别名解析。

3.3　分类特征表示

基于第 2 节中对接口 IP 的时延、路径、Whois、对应主机等信息的统计分析，本文给出了用于对别名 IP 和非别名 IP 进行分类的四维特征：时延相似度、路径相似度、Whois 信息相似度和主机信息相似度。之所以设计这样四维特征并用于分类，是因为尽管大量的统计分析表明，别名 IP 与非别名 IP 在时延、路径等多种数据方面存在差异，所获取的各类数据仍可能会受到如时延膨胀、探测路径的完整性、Whois 的更新频率以及主机的监测时长等不同因素的影响，但这些因素将仅仅影响到相关的单维特征的分类效果；四维特征间的相关性较弱，不会相互影响。因此，在避免特征冗余的情况下，通过将多维特征用于分类，使在个别单维特征受到影响时，最终依然有望获取到相对较好的分类效果。此外，所设计的四维特征对于分类是互补的，通过时延、路径相似度可准确识别出地理分布不临近的非别名接口 IP，在此基础上，Whois、主机信息相似度能够进一步将别名 IP 和非别名 IP 区别开。本节对特征的表示进行具体介绍。

3.3.1　时延相似度

由 2.1 节的统计分析可知，同一源 IP 到存在别名关系的两个 IP 的时延往往较为相似，到不存在别名关系的两个 IP 的时延相似度较低，但受实际网络状况对时延的影响，仍有个例不符合该规律。仅利用单一源 IP 到任意两个 IP 的时延相似度，难以判别 IP 间是否存在别名关系。而从多个源 IP 分别获取到两个 IP 的时延相似度，能够减少网络状况的影响。为此，对待判别的 IP 对，采取从多个不同位置的源 IP，分别获取到两个 IP 的时延。对于其中的每个 IP，利用获取到的多个时延，为该 IP 构造时延向量。对待判别的两个 IP，计算其时延向量的相似度，并作为一维分类特征，具体如下。

设任意两个待解析的 IP 为 IP_i, IP_j，分布式部署 n 个位于不同位置的探测源 N_1~N_n，从每个探测源分别对这两个 IP 进行多次探测，对每个 IP 获取一个最小时延，以尽可能减小网络拥塞等影响。对于 IP_i，其 n 个最小时延为 $t_{i,1}, t_{i,2}, \cdots, t_{i,n}$，对于 IP_j，其 n 个最小时延为 $t_{j,1}, t_{j,2}, \cdots, t_{j,n}$。为 IP_i, IP_j 构造时延向量 $\boldsymbol{D}_i(t_{i,1}, t_{i,2}, \cdots, t_{i,k}, \cdots, t_{i,n})$, $\boldsymbol{D}_j(t_{j,1}, t_{j,2}, \cdots, t_{j,k}, \cdots, t_{j,n})$。利用式(1)计算 D_i 与 D_j 的相似度 S_d，将其作为特征值。

$$S_d=\sqrt{\sum_{k=1}^{n}(t_{i,k}-t_{j,k})^2},\ 1\leqslant k\leqslant n \tag{1}$$

3.3.2　路径相似度

由 2.2 节的分析可知，一定时间内，路由转发的下一跳往往是不变的，从源 IP 到目的 IP 的路径相对固定。存在别名关系的接口 IP 处于同一路由器上，当从同一探测源对其进行探测时，探测路径（方向和长度）往往较为相似。对待解析的 IP 对，分别获取不同源 IP 到两个接口 IP 的路径，并根据路径构造向量，从而计算两个 IP 的路径相似度，作为分类特征。

设任意两个待解析的接口 IP 为 IP_i, IP_j，从探测源 N_1~N_n 分别对该两个 IP 进行 m 次探测。由于路由器至少拥有两个接口，一些大型核心骨干路由器通常拥有 10~30 个接口[7]，为保证能够尽可能全地发现探测路径上每一跳路由器的接口 IP，应置探测次数 m 大于路由器接口数量，如取 m=50。本文按如下方式计算从探测源 N_n 到 IP_i, IP_j 的路径相似度。

对于两个 IP 的探测路径，分别取 m 次探测

中出现次数最多的路径跳数作为从探测源 N_n 到该 IP 的探测路径长度，将从 N_n 得到的 IP_i, IP_j 的路径向量分别记为 $path_{i,n}$，$path_{j,n}$，$path_{i,n}$ 表示为 $(A_{1,n}, A_{2,n}, \cdots, A_{l,n}, \cdots, A_{x,n})$，$path_{j,n}$ 表示为：$(B_{1,n}, B_{2,n}, \cdots, B_{l,n}, \cdots, B_{y,n})$。其中，$x$，$y$ 分别为 IP_i, IP_j 路径的长度，$A_{l,n}$，$B_{l,n}$ 分别为两个 IP 路径上第 l 跳出现的所有路由器接口 IP 构成的集合。若 IP_i, IP_j 为别名 IP，则应有 $A_{l,n} \approx B_{l,n}$，$(A_{l,n} \cap B_{l,n}) \approx (A_{l,n} \cup B_{l,n})$，且 $x \approx y$；若 IP_i, IP_j 为非别名 IP，则 $A_{l,n}$ 与 $B_{l,n}$，x 与 y 有一定差异。因此，将从 N_n 得到的 IP_i, IP_j 的路径的相似度 S_n 表示为

$$S_n = \sqrt{\sum_{l=1}^{\mathrm{MAX}(x,y)} \left(\frac{|A_{l,n} \cap B_{l,n}|}{|A_{l,n} \cup B_{l,n}|} \right)^2} \tag{2}$$

式(2)中，当 $x<y$ 时，置 $A_{x+1,n} \sim A_{y,n}$ 为 Ø；反之，当 $y<x$ 时，置 $B_{y+1,n} \sim B_{x,n}$ 为 Ø。最终，IP_i, IP_j 的路径相似度 S_p 可表示为从 n 个探测源获取的路径相似度的平均值，即

$$S_p = \frac{\sum_{r=1}^{n} S_r}{n},\ 1 \leqslant r \leqslant n \tag{3}$$

3.3.3 Who is 信息相似度

通过 2.3 节关于 IP 的 Whois 信息分析可知，对于大多数存在别名关系的一对 IP，其 Whois 信息较为一致，但统计发现，少量不存在别名关系的 IP，其个别 Whois 信息项相同，这可能是由于信息更新不及时等导致。为了更好地根据 Whois 信息相似程度判断 IP 间是否存在别名关系，对不同的 Whois 信息项赋权值，计算 IP 间 Whois 信息的相似度，并将其作为一维分类特征，具体表示如下。

存在别名关系的两个 IP，当其 Whois 信息完全相同时，总条数记为 H，记第 h 条 Whois 信息为 I_h，$1 \leqslant h \leqslant H$。设任意两个待解析的 IP 为 IP_i, IP_j，当其第 h 条信息相同时，有 $v(I_h)=1$，否则 $v(I_h)=0$。

一些不存在别名关系的 IP，个别 Whois 信息项（如所属网段、网络名称、所属国家、状态信息等）可能相同。这几项信息对于判别 IP 间是否存在别名关系的贡献，小于仅当 IP 间存在别名关系时才会相同的 Whois 信息，因此本文为不同信息项赋不同的权值。设该 4 条信息项构成的集合为 I，则将信息项 I_h 的权值 $w(I_h)$表示为

$$w(I_h) = \begin{cases} p, I_h \in I \\ q, I_h \notin I \end{cases} \tag{4}$$

其中，$p<0.5<q$，本文取 $p=0.1$，$q=0.9$。对于 IP_i 与 IP_j，设其相同信息项构成集合为 K，则其 Whois 信息相似度 S_w 可表示为

$$S_w = \frac{\sum_{g=1}^{|K|} w(I_g) \cdot v(I_g)}{\sum_{h=1}^{H} w(I_h) \cdot v(I_h)} = \frac{\sum_{g=1}^{|K|} w(I_g)}{\sum_{h=1}^{H} w(I_h)},\ 1 \leqslant g, h \leqslant |K| \leqslant H \tag{5}$$

3.3.4 主机信息相似度

根据 2.4 节中大量探测数据的统计分析可知，存在别名关系的 IP 对应的主机，在操作系统版本、端口开放情况以及主机运行状态方面，较为一致，尤其在主机运行状态和端口开放方面，具有高度的一致性。不存在别名关系的 IP，其对应主机的上述信息，往往不同，但个别 IP 的操作系统版本或部分开放端口相同。因此，为了充分考虑不同主机信息的特点，更好地依据主机信息对 IP 间别名关系进行判断，按如下方式计算 IP 对应主机的信息相似度。

设任意两个待解析的 IP 为 IP_i, IP_j，从 n 个探测源分别对其进行 Z 次探测，并根据每一次的探测结果，获取 IP 对应主机的操作系统版本、端口开放情况以及主机运行状态信息。对于任意时刻，只有在 IP_i, IP_j 对应的路由器主机的运行状态完全相同的情况下，这两个 IP 才有可能配置在同一路由器不同端口上，即存在别名关系。所以，在确保 IP_i, IP_j 对应的主机运行状态相同的情况下，根据主机操作系统版本、开放端口数量及端口状态，计算两个 IP 对应的主机信息相似度如下。

第 z 次探测时，对于 IP_i, IP_j，若其对应路由器主机操作系统相同，则系统相似度 $s_{1,z}=1$，否则 $s_{1,z}=0$。设 IP_i, IP_j 对应路由器主机开放相同的端口数量为 n_{same}，总开放端口数量为 n_{all}，则端口开放相似度 $s_{2,z}=\frac{n_{\mathrm{same}}}{n_{\mathrm{all}}}$。对于第 z 次探测时，IP_i, IP_j 的主机信息相似度 S_z 可表示为 $as_{1,z}+bs_{2,z}$，其中，a, b 为权值。大量的统计分析发现，个别不存在别名

关系的 IP，其对应主机的操作系统存在相同的情况，但开放的端口数量及状态几乎没有相同的，相对来说，端口开放相似度对于判别 IP 间的别名关系贡献更大，因此在计算 S_z 时，有 $a<0.5<b$，本文取 $a=0.3, b=0.7$。

考虑到别名 IP 间，上述信息任意时刻较为相似，而非别名 IP 则不然，因此，将 IP_i, IP_j 的主机信息相似度 Sh 表示为所有探测中信息相似度的均值，即

$$S_h = \frac{\sum_{z=1}^{nZ} S_z}{nZ} \tag{6}$$

4　实验设计及结果

为了验证所提算法 MLAR 的有效性，本节给出了多组测试及结果分析。4.1 节介绍了样本数据的来源以及实验相关的设置；4.2 节分别对算法中的非别名 IP 过滤规则，以及别名解析算法的效果进行测试；4.3 节采用几种不同的方法进行多组别名解析，并从正确率、效率及应用于 IP 定位的效果等方面，对不同方法进行对比分析。

4.1　实验设置

实验中接口 IP 样本数据主要来源于 CAIDA。该网站提供了大量的可靠路由器级网络拓扑数据，其中包含路由节点，以及节点的接口 IP 和位置信息，每个节点的多个接口 IP 相互间存在别名关系，通过将同一路由节点的不同接口 IP 进行两两组合，可构造别名 IP 集；同时，不同节点间的接口 IP，不存在别名关系，将不同路由节点的接口 IP 进行两两组合，构造非别名 IP 集。

为获取丰富的时延、路径等探测数据，需在待解析的接口 IP 周围分散部署多个探测源。对于上述样本中属于中国的路由节点接口 IP，在郑州、北京、上海、广州、天津、成都等地部署 10 个探测源，并从每个探测源对各个 IP 进行探测；同样地，对于属于美国的路由节点接口 IP，在纽约、芝加哥、亚特兰大、华盛顿、迈阿密、西雅图等地部署 10 个探测源，并从每个探测源对各个 IP 进行探测。文献[34]指出，网络的路由路径在短时间内（如一个月）是相对稳定的。通过对大量探测数据的统计分析后发现，路由路径的确存在上述稳定性。因此，为保证探测数据的可靠性，应保证在较短的时间周期内对接口 IP 进行探测。接口 IP 的 ISP、Whois 信息，主要通过查询相关 IP 信息库获取，IP 对应的路由器主机信息，则利用 Nmap 获取。

利用获取到的 IP 的时延、路径等信息，依据 3.1 节中别名解析算法的具体步骤，对样本进行如下的别名解析测试。具体的实验设置如表 4 所示。

4.2　别名解析测试及结果分析

本节利用已知样本，分别对 MLAR 的非别名 IP 过滤效果及别名解析效果进行测试，并分析测试结果。

4.2.1　非别名 IP 过滤测试

MLAR 给出了用于非别名 IP 过滤的规则，为了验证所设计规则的有效性，利用如 4.1 节所述的样本，在获取到所需相应数据后，本节利用这些规则进行过滤测试。表 5 给出了对分布于中国北京、上海及美国纽约、迈阿密的样本的过滤结果，其中测试时 4 个城市的别名 IP 及非别名 IP 数量均为 1×10^6 对。

分析表 5 结果可得，仅有个别别名 IP 对被所设计的规则当作非别名 IP 对过滤掉，其中有 41 对位于中国上海的 IP 被规则 3)过滤掉，有 23 对位于美国迈阿密的 IP 被规则 4)过滤掉，被过滤掉的主要原因是一对 IP 中的其中一个 IP 可能由于分组丢失等原因导致探测不通，而另一个探测可达，该情况极少出现；通过规则 1)~5)，4 个城市中分别有 83.4%、81.7%、84.6%、86.2%的非别

表 4　实验设置
Table 4　Experimental settings

样本来源	样本分布	样本数量			探测源分布
		接口 IP	别名 IP 对	非别名 IP 对	
CAIDA	中国	9.153×10^6	5.729×10^7	7.113×10^9	郑州、北京、上海、广州、杭州、天津、成都、西安、济南、武汉
	美国	1.075×10^7	8.211×10^7	3.025×10^{11}	纽约、芝加哥、亚特兰大、华盛顿、迈阿密、西雅图、洛杉矶、达拉斯、旧金山、菲尼克斯

表 5 过滤结果

Table 5 The filtering results

过滤规则	通过各规则，不同地区被过滤掉的两类样本比例							
	别名 IP				非别名 IP			
	北京	上海	纽约	迈阿密	北京	上海	纽约	迈阿密
1)	0	0	0	0	22.2%	18.9%	24.5%	26.3%
2)	0	0	0	0	5.3%	3.6%	7.1%	5.0%
3)	0	0.0041%	0	0	26.6%	30.6%	26.9%	28.9%
4)	0	0	0	0.0023%	20.4%	21.1%	22.6%	23.3%
5)	0	0	0	0	8.9%	7.5%	3.5%	2.7%
合计	**0**	**0.0041%**	**0**	**0.0023%**	**83.4%**	**81.7%**	**84.6%**	**86.2%**

名 IP 对被准确过滤掉。由此可以看出，MLAR 给出的过滤规则能够准确过滤掉大部分非别名 IP 对，同时保留别名 IP 对，使用该规则，能够大大减少别名解析的工作量，从而提高效率。

4.2.2 别名解析测试

为了测试 MLAR 的别名解析效果，从 4.1 节所述的样本中取别名 IP 对及非别名 IP 对，分别构造集合 S_0，S_1，其中，分布于中国的样本数量为 2×10^7，分布于美国的样本数量为 3×10^7。本文采用 SVM（support vector machine）分类器，对样本进行分类测试。为了考察 MLAR 对样本数据量的依赖程度，分别设置如下 3 组测试，每组测试进行 3 次随机抽样：① $\frac{1}{4}$ 的数据用于构造训练集，$\frac{3}{4}$ 的数据用于构造测试集，对应 a_1~a_3；② $\frac{1}{2}$ 的数据用于构造训练集，$\frac{1}{2}$ 的数据用于构造测试集，对应 b_1~b_3；③ $\frac{3}{4}$ 的数据用于构造训练集，$\frac{1}{4}$ 的数据用于构造测试集，对应 c_1~c_3。测试结果如表 6 所示。其中，分类正确率（Acc，Accuracy）、漏报率（Ma，Missing alarm）、虚警率（Fa，False alarm）计算如下，Acc 为 $\frac{\mathrm{TT}+\mathrm{FF}}{X+Y}$，Ma 为 $\frac{\mathrm{TF}}{Y}$，Fa 为 $\frac{\mathrm{FT}}{X}$。X 表示测试集中别名 IP 对的数量，Y 表示测试集中非别名 IP 对的数量，TT 表示被正

表 6 训练、测试集构造及对应分类结果

Table 6 Training set, test set construction and corresponding classification results

测试序号	训练集		测试集		分类结果		
	别名 IP 对	非别名 IP 对	别名 IP 对	非别名 IP 对	Acc	Ma	Fa
a_1	$\frac{1}{4}$ S_0	$\frac{1}{4}$ S_1	$\frac{3}{4}$ S_0	$\frac{3}{4}$ S_1	96.3%	2.7%	4.7%
a_2	$\frac{1}{4}$ S_0	$\frac{1}{4}$ S_1	$\frac{3}{4}$ S_0	$\frac{3}{4}$ S_1	95.4%	4.4%	4.8%
a_3	$\frac{1}{4}$ S_0	$\frac{1}{4}$ S_1	$\frac{3}{4}$ S_0	$\frac{3}{4}$ S_1	96.0%	3.8%	4.1%
b_1	$\frac{1}{2}$ S_0	$\frac{1}{2}$ S_1	$\frac{1}{2}$ S_0	$\frac{1}{2}$ S_1	96.4%	3.4%	3.7%
b_2	$\frac{1}{2}$ S_0	$\frac{1}{2}$ S_1	$\frac{1}{2}$ S_0	$\frac{1}{2}$ S_1	95.9%	4.0%	4.2%
b_3	$\frac{1}{2}$ S_0	$\frac{1}{2}$ S_1	$\frac{1}{2}$ S_0	$\frac{1}{2}$ S_1	96.8%	3.1%	3.3%
c_1	$\frac{3}{4}$ S_0	$\frac{3}{4}$ S_1	$\frac{1}{4}$ S_0	$\frac{1}{4}$ S_1	96.0%	3.8%	4.2%
c_2	$\frac{3}{4}$ S_0	$\frac{3}{4}$ S_1	$\frac{1}{4}$ S_0	$\frac{1}{4}$ S_1	96.9%	2.9%	3.2%
c_3	$\frac{3}{4}$ S_0	$\frac{3}{4}$ S_1	$\frac{1}{4}$ S_0	$\frac{1}{4}$ S_1	96.6%	3.3%	3.5%

确分类的别名 IP 对的数量，FF 表示被正确分类的非别名 IP 对的数量，FT 表示被错误分类的别名 IP 对的数量，TF 表示被错误分类的非别名 IP 对的数量。

由表 6 可得，总体而言，MLAR 所获得的正确率较高，漏报率和虚警率较低。上述 3 组共 9 次测试的正确率维持在 95%~97%，测试 a_1~a_3 的平均正确率为 95.9%，b_1~b_3 的平均正确率约为 96.4%，c_1~c_3 的平均正确率为 96.5%。由相同样本量的测试结果可得，MLAR 的性能具有一定的稳定性。对比测试 a_1~a_3，b_1~b_3 与 c_1~c_3 可以看出，即使使用少量训练样本数据，也能获得相对较好的分类模型及分类效果。

4.2.3　不同特征组合的分类效果测试

为了进一步验证各维特征对分类的贡献及其别名解析效果，本节分别利用不同的特征组合，对 4.2 节所述的别名 IP 样本集合 S_0 及非别名 IP 样本集合 S_1 进行分类测试实验，$\frac{3}{4}$ 的数据用于构造训练集，$\frac{1}{4}$ 的数据用于构造测试集，每组测试均采用相同的样本。不同特征组合的分类效果如表 7 所示。

表 7　不同特征组合的分类效果

Table 7　Classification results of different feature combinations

测试序号	特征组合	分类效果		
		Acc	Ma	Fa
1	F_1	81.1%	18.7%	19.0%
2	F_2	84.9%	14.9%	15.2%
3	F_3	82.4%	17.5%	17.8%
4	F_4	85.3%	14.3%	15.0%
5	F_1、F_2	89.1%	10.0%	11.7%
6	F_1、F_3	88.6%	10.9%	11.8%
7	F_1、F_4	90.5%	9.0%	10.0%
8	F_2、F_3	87.7%	12.0%	12.5%
9	F_2、F_4	92.3%	7.7%	7.8%
10	F_3、F_4	91.5%	7.8%	9.1%
11	F_1、F_2、F_3	93.9%	6.0%	6.2%
12	F_1、F_2、F_4	94.4%	5.1%	6.1%
13	F_1、F_3、F_4	94.3%	5.3%	6.0%
14	F_2、F_3、F_4	95.2%	4.5%	5.2%
15	F_1、F_2、F_3、F_4	96.2%	3.6%	4.0%

由表 7 可得，在使用相同样本时，利用不同特征组合的分类效果不同，单维特征的分类效果不如多维特征组合的分类效果，采用特征维数越多，分类效果越好。单维特征对分类的贡献由高到低依次为：主机信息相似度 F_4、路径相似度 F_2、Whois 信息相似度 F_3、时延相似度 F_1。这主要是由于一段时间内 IP 对应主机信息、探测路径信息等相对稳定可靠，在这些信息方面别名 IP 与非别名 IP 的差异明显，而少量 IP 的 Whois 信息更新不及时，网络拥塞等导致时延测量不够准确，使在这两类信息方面别名 IP 与非别名 IP 的差异相对弱一些。但由于存在部分非别名 IP 对应主机信息等高度相似，而其时延相似度可能有较大差异，或路径信息相似而 Whois 信息差异明显等情况，此时采用单维特征难以对别名 IP 及非别名 IP 进行分类，采用四维特征的组合进行分类效果更佳，这说明所设计的四维特征不是冗余的。

4.2.4　不同分类算法的分类效果测试

为了验证所构建的四维特征的分类效果，以及测试不同数据集下分类模型的稳健性，本节设计两组测试：对于第一组测试，从 4.2 节中别名 IP 及非别名 IP 样本集合 S_0、S_1 中随机抽取 $\frac{3}{4}$ 的数据用于构造训练集，$\frac{1}{4}$ 的数据用于构造测试集；对于第二组测试，采用国内 ISP 提供的一些路由器接口 IP（包括 21 368 对别名 IP 以及 25 016 对非别名 IP）作为实验数据，同样从别名 IP 及非别名 IP 样本中随机抽取 $\frac{3}{4}$ 的数据用于构造训练集，$\frac{1}{4}$ 的数据用于构造测试集。分别利用支持向量机、逻辑回归（LR，logistic regression）、朴素贝叶斯分类器（NBC，naive Bayesian classifier）进行分类测试实验，对于不同数据集，每种分类模型进行 3 次测试，分类效果如表 8 所示。

由表 8 可得，使用所构建的四维特征，采用 SVM、LR、NBC 这 3 种不同的分类模型，对不同种类数据集的分类正确率都较高。其中，对于 CAIDA 数据集，3 种分类器所得平均正确率分别为 96.4%、95.7%、95.5%，对于 ISP 数据集，3 种分类器所得平均正确率分别约为 97.2%、96.8%、97.1%，说明 MLAR 所构建的分类特征能够较为

可靠地区分别名 IP 及非别名 IP。使用不同分类模型所得分类结果，漏报率、虚警率都较低，对于 CAIDA 数据集，平均漏报率分别约为 3.5%、4.2%、4.4%，平均虚警率分别约为 3.8%、4.4%、4.7%。对于 ISP 数据集，平均漏报率分别约为 2.6%、3.2%、2.7%，平均虚警率分别约为 3.0%、3.3%、3.2%，漏报率都低于虚警率，说明通过 MLAR 所构建的特征将非别名 IP 判为别名 IP 的可能性很小，能够从路由器接口 IP 中准确识别出别名 IP。

表 8 不同分类算法的分类效果

Table 8 Classification results of different algorithms

数据集来源	测试序号	分类模型	分类效果		
			Acc	Ma	Fa
CAIDA	1	SVM	96.5%	3.3%	3.6%
	2		96.2%	3.7%	4.0%
	3		96.4%	3.4%	3.7%
	4	LR	96.0%	3.8%	4.2%
	5		95.3%	4.7%	4.6%
	6		95.7%	4.2%	4.4%
	7	NBC	95.5%	4.3%	4.6%
	8		94.9%	5.0%	5.3%
	9		96.0%	3.8%	4.1%
ISP	10	SVM	97.1%	2.6%	3.3%
	11		97.5%	2.3%	2.8%
	12		97.1%	2.8%	3.0%
	13	LR	96.8%	3.2%	3.2%
	14		96.9%	2.8%	3.4%
	15		96.6%	3.5%	3.4%
	16	NBC	97.2%	2.6%	2.9%
	17		97.3%	2.4%	3.1%
	18		96.7%	3.0%	3.6%

4.3 与现有典型方法的比较

准确、高效的别名解析，对于获取能够反映真实网络结构的路由器级网络拓扑，从而支撑 IP 定位意义重大。现有部分典型方法如 RadarGun[19]、MIDAR[20]、TreeNET[24]等，在别名解析方面具有相对良好的效果。本节从别名解析的正确率和效率，以及对 IP 定位的帮助等方面，对这几种方法与 MLAR 进行测试及对比 分析。

4.3.1 别名解析准确性对比

为了验证所提别名解析算法 MLAR 的准确性，从 4.1 节所述的样本中取别名 IP 对及非别名 IP 对，其中分布于中国的样本数量为 3×10^7 个，分布于美国的样本数量为 5×10^7 个。分别利用 RadarGun、MIDAR、TreeNET 进行 5 次别名解析；对于 MLAR，使用 4.2.2 节测试 c_1 中训练好的分类模型进行 5 次别名解析。表 9 给出了不同方法每一次测试对应的正确率、漏报率及虚警率。

由表 9 可以看出，RadarGun、MIDAR、TreeNET 及 MLAR 都能够获得一定的效果，平均的正确率分别约为 82.7%、91.4%、90.6%、95.8%，相对而言，MIDAR、TreeNET 和 MLAR 的正确率较高。MLAR 相比前 3 种方法正确率分别平均提高了 15.8%、4.8%、5.7%。上述测试结果中，4 种方法的 5 次测试所得正确率的标准差分别为 0.038 0、0.012 9、0.006 2、0.005 6，相比其他两种方法，TreeNET 及 MLAR 多次测试结果正确率较为一致，具有一定的稳定性。此外，在测试时，将别名 IP 对作为正例样本，非别名 IP 对作为负例样本，结合 4.2.2 节的测试结果可得，对 MLAR 多次测试所得漏报率都低于虚警率，说明虽然 MLAR 存在将部分别名 IP 对误判为非别名 IP 对的情况，但通过 MLAR 所获取的别名 IP 对较为

表 9 不同方法多次测试结果对比

Table 9 Comparison of test results of different methods

测试序号	RadarGun			MIDAR			TreeNET			MLAR		
	Acc	Ma	Fa	Acc	Ma	Fa	Acc	Ma	Fa	Acc	Ma	Fa
1	84.9%	17.2%	13.0%	92.5%	7.4%	7.5%	90.6%	9.9%	9.0%	**96.7%**	**2.9%**	**3.7%**
2	80.4%	18.8%	20.4%	91.6%	8.5%	8.3%	89.7%	9.9%	10.7%	**95.3%**	**4.3%**	**5.1%**
3	76.6%	23.9%	22.8%	90.2%	9.8%	9.8%	90.1%	10.1%	9.8%	**96.1%**	**3.5%**	**4.3%**
4	84.1%	15.2%	16.6%	89.8%	9.9%	10.6%	91.3%	8.6%	8.8%	**95.6%**	**4.8%**	**3.9%**
5	87.5%	13.4%	11.5%	93.1%	7.2%	6.6%	91.2%	8.0%	9.6%	**95.2%**	**4.6%**	**5.0%**

准确可靠，这对于 IP 定位尤为重要。

4.3.2　别名解析效率对比

为了验证 MLAR 对别名解析的高效性，同样采用 4.1 节所述的样本数据，分别利用 RadarGun、MIDAR、TreeNET 及 MLAR，对不同规模的网络（即包含不同数量的接口 IP），在相同的实验环境下，分别进行 3 次测试，并对测试所用时长进行对比分析。考虑到 MIDAR 需分布式多源探测以提高效率，且 MLAR 需要通过多个探测源获取相关数据，为了公平比较不同方法的效率，对于 MIDAR 及 MLAR，通过相同配置的 10 台主机配合完成测试，而对于 RadarGun 及 TreeNET，则将每一次测试的样本平均分为 10 份，分别在上述 10 台主机上利用这两种方法进行别名解析，并记录 10 台主机中的最长耗时。

当网络规模不断增大，接口 IP 数量由 1×10^6 个递增到 5×10^6 个时，别名 IP 对的数量分别为 7.325×10^6，1.901×10^7，2.421×10^7，2.944×10^7，3.602×10^7；非别名 IP 对的数量分别为 7.903×10^8，1.311×10^9，2.404×10^9，3.224×10^9，4.003×10^9。对于 RadarGun 和 TreeNET，对所有的 IP 对都要进行解析；MIDAR 认为当从两个目标 IP 获得的 IP-ID 序列变化速率相似度低时，这两个 IP 不可能共享 IP-ID 计数器，即不存在别名关系，依据该理论可过滤掉的 IP 对的比例分别为 10.6%，15.3%，25.1%，19.9%，22.10%；而对于 MLAR，通过规则过滤掉的不存在别名关系的 IP 对的比例分别为 67.3%，72.7%，71.4%，69.9%，75.2%。图 5 给出了不同方法需要进行解析的样本数量，由图 5 可以看出，MLAR 需要解析的样本数量最少。

表 10 及图 6 给出了随着网络规模的增大，接口 IP 数量的增加，不同方法 3 次测试所用时长。

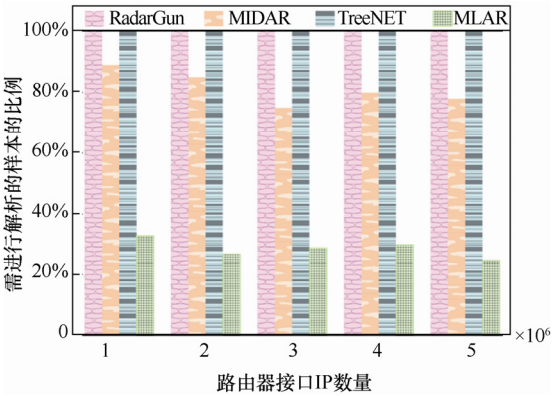

图 5　不同方法需要进行解析的样本的比例

Figure 5　The proportion of aliases that need to be resolved by different methods

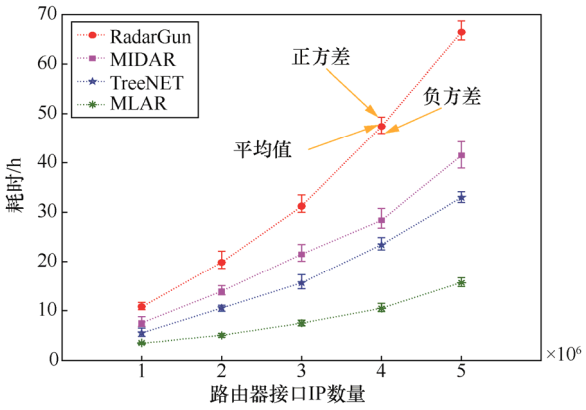

图 6　不同方法耗时

Figure 6　Duration of different methods

根据表 10 及图 6 的结果可以看出，接口 IP 数量不同，各个方法所用时长不同，且随着 IP 数量的增加，所用时长都在增加，每次测试 RadarGun 耗时最长，其次为 MIDAR、TreeNET，MLAR。由图中曲线的斜率变化可以看出，相比 MLAR，RadarGun、MIDAR、TreeNET 所用时长增长的速率较大，当接口 IP 数量为 1×10^6 个时，RadarGun、MIDAR、TreeNET 的平均耗时分别约

表 10　不同方法效率对比

Table 10　Efficiency comparison of different methods

方法	随着接口 IP 数量递增，不同方法 3 次测试所用时长/h														
	1×10^6			2×10^6			3×10^6			4×10^6			5×10^6		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
RadarGun	10.4	10.2	11.7	22.1	18.9	18.6	30.1	33.5	30.0	46.9	46.0	49.3	65.7	64.9	68.7
MIDAR	7.0	8.8	6.8	13.2	15.1	13.5	20.1	23.5	20.9	27.5	26.8	30.8	41.1	44.3	38.9
TreeNET	4.8	5.2	6.5	9.9	10.6	11.1	14.4	15.2	17.3	23.0	22.4	24.9	31.9	32.8	34.1
MLAR	**3.5**	**3.3**	**3.6**	**5.4**	**5.1**	**4.6**	**8.1**	**7.4**	**6.9**	**11.5**	**9.8**	**10.2**	**15.5**	**16.7**	**14.9**

为 MLAR 的 3.1 倍，2.2 倍，1.6 倍，但当接口 IP 数量增加到 5×10^6 时，分别增加到了 4.2 倍，2.6 倍，2.1 倍。这是为了能够获取到 IP-ID，RadarGun 和 MIDAR 需要对每个 IP 进行大量探测，但 MIDAR 在别名解析前进行了初步的过滤，而 RadarGun 没有使用任何过滤机制，因此 MIDAR 效率高。TreeNET 没有设定过滤规则，但其根据 IP 对探测的响应情况，不完全依赖于 IP-ID，还综合了基于路由器主机名的解析等方法，而这种无须探测的解析效率极高，因此 TreeNET 总体效率高于 MIDAR。对于 MLAR，在别名解析前，利用多个探测源的探测结果，平均过滤掉了 71.3%的非别名 IP 对。本文对 IP 对应主机运行状态监测时长设定为 2.5 h，在监测的同时，并行获取用于别名解析的时延、探测路径等数据，可节省大量时间，效率最高，且仅当需要解析的 IP 数量较大时，耗时才出现明显变化。

此外，曲线上“I”形的上端和下端分别表示耗时的正方差及负方差值，线上的点表示耗时的均值，通过对比 4 条曲线可以看出，对于相同接口 IP 数量的多次测试，RadarGun、MIDAR 耗时最不稳定，差异较大，而 MLAR 耗时相对稳定。上述结果在一定程度上说明 MLAR 在别名解析效率方面有一定优势。

4.3.3 应用于 IP 定位的效果对比

为了进一步验证所提别名解析算法 MLAR 的有效性，本节对上述几种方法在实际 IP 定位中的应用效果进行对比。

文献[4]提出 SLG——一种逐层逼近的街道级定位方法，并在最后一层，将与目标 IP 存在最近共同路由器且相对时延最小的地标的位置，作为目标的位置估计。由于探测获取的拓扑实际为路由器接口级拓扑，当地标与目标分别与最近共同路由器的不同接口 IP 相连时，如果不进行别名解析，则无法得知二者实际与同一路由器相连，因此无法通过地标的位置估计目标 IP 的位置，定位将失败。文献[5]寻找与目标 IP 存在最近共同路由器且相对时延最小的 3 个地标，并根据三点定位思想对目标 IP 进行街道级定位，与 SLG 面临的问题类似，该算法的前提条件也是找到最近共同路由器，因此若想降低定位失败率，在寻找共同路由器前需要进行别名解析。文献[6]则利用划分的 PoP 对目标 IP 进行城市级定位，该方法需要通过别名解析，将城市内部本应属于同一个大规模 PoP 的多个小 PoP 进行合并，提高 PoP 的完整性，并用于 IP 定位。因此，别名解析的效果将一定程度上决定所获取 PoP 的完整性，从而决定 IP 定位的效果。

将不同的别名解析方法运用到上述 3 种典型的定位方法中，对实际网络环境中的目标 IP 进行定位测试，并对定位结果进行分析。对于 SLG 与 LENCR，分别在中国北京、美国加利福尼亚州取 1 000、3 000 个街道级地标作为待定位目标 IP，对于 PoPG，分别在中国北京、美国加利福尼亚州取 50 000 个城市级地标作为待定位目标 IP，分别对 3 种方法在使用及不使用别名解析时对目标 IP 的定位效果进行对比，表 11 给出了具体的定位结果。

表 11 给出了在使用及不使用别名解析两种情况下，3 种定位算法对中国北京及美国加利福尼亚州的目标 IP 进行定位的失败率。其中，每种定位算法下的数据表示在定位过程中，该定位算

表 11 定位测试结果对比

Table 11 Comparison of geolocation test results

别名解析方法	在使用不同别名解析方法前后，不同定位算法对两个区域的目标 IP 的定位失败率					
	北京（中国）			加利福尼亚州（美国）		
	SLG[4]	LENCR[5]	PoPG[6]	SLG[4]	LENCR[5]	PoPG[6]
无	28.9%	31.6%	18.7%	25.1%	26.5%	15.2%
RadarGun	21.6%	24.4%	13.2%	19.3%	16.4%	11.9%
MIDAR	13.5%	16.3%	9.6%	15.9%	13.7%	9.7%
TreeNET	16.1%	19.2%	11.8%	16.2%	14.6%	10.4%
MLAR	**9.0%**	**10.4%**	**6.7%**	**9.5%**	**10.3%**	**7.3%**

法使用对应的别名解析方法后，对目标 IP 定位的失败率，当别名解析方法为无时，表示在该定位算法的定位过程中，不使用任何的别名解析方法。由表 11 可以得出，在使用别名解析方法前后，3 种定位算法对两个地区的目标 IP 的定位效果差别较大，使用别名解析后，定位失败率明显降低。其中，对于 SLG，相比未使用别名解析，使用 RadarGun、MIDAR、TreeNET 及 MLAR 后定位失败率平均分别降低了 24.2%，45.0%，39.9%，65.5%；对于 LENCR，分别平均降低了 30.4%，48.4%，42.1%，64.1%；对于 PoPG，分别平均降低了 25.6%，42.4%，34.2%，58.1%。通过对比发现，使用所提别名解析算法 MLAR 后，3 种定位算法的定位失败率降低最多，间接说明了 MLAR 的别名解析效果最好。

5　结束语

现有一些典型的别名解析方法所需数据难以获取，别名解析准确率难以保证，在解析前未对大量不可能存在别名关系的 IP 对过滤，别名解析的效率低，导致这些方法难以满足大规模网络的别名解析需求，难以支撑 IP 定位等实际应用。为此，本文提出了一种面向 IP 定位的大规模网络别名解析算法 MLAR。MLAR 利用接口 IP 较易于获取时延、路径等相关数据，并基于目标区域内别名 IP 与非别名 IP 在这些数据方面的统计差异，排除大量的非别名 IP；利用机器学习对区域内剩余 IP 对进行别名解析。结合 MLAR，本文准确地刻画大规模网络中路由节点连接及拓扑，从而降低基于路由节点连接关系的 IP 定位方法的失败率。本文采用 CAIDA 提供的分布于中国和美国一些城市的百万级样本数据对 MLAR 进行了测试实验。结果表明与现有的 RadarGun、MIDAR、TreeNET 等典型方法相比，MLAR 的正确率、效率更高，更适用于大规模网络，能够更好地帮助 IP 定位。但针对特定目标区域进行别名解析时，所提算法对区域内已知样本的数量仍有一定的要求。此外，网络拥塞、路由变化、一些数据（如 Whois）的更新频率等因素会影响算法的效果。

参考文献：

[1] CANBAZ M A. Internet topology mining: from big data to network science[D]. Reno: University of Nevada, 2018.

[2] KARDES H, GUNES M H, SARAC K. Graph based induction of unresponsive routers in internet topologies[J]. Computer Networks, 2015, 81: 178-200.

[3] COSKUN I E, CANBAZ M A, GUNES M H. Efficient AS network topology measurement based on ingress to subnet reachability[C]// IEEE 41st Conference on Local Computer Networks Workshops. 2016: 87-95.

[4] WANG Y, BURGENER D, FLORES M, et al. Towards street-level client-independent IP geolocation[C]//Symposium on Network System Design and Implementation. 2011: 27-27.

[5] CHEN J, LIU F, SHI Y, et al. Towards IP location estimation using the nearest common router[J]. Journal of Internet Technology, 2018, 19(7): 2097-2110.

[6] YUAN F, LIU F, HUANG D, et al. A high completeness PoP partition algorithm for IP geolocation[J]. IEEE Access, 2019, 7: 28340-28355.

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

[8] MARCHETTA P, PESCAPÉ A. DRAGO: detecting, quantifying and locating hidden routers in traceroute IP paths[C]// Proceedings IEEE International Conference on Computer Communications. 2013: 3237-3242.

[9] LI R, SUN Y, HU J, et al. Street-level landmark evaluation based on nearest routers[J]. Security and Communication Networks, 2018(2): 1-12.

[10] HINGANT J, ZAMBRANO M, PÉREZ F J, et al. HYBINT: a hybrid intelligence system for critical infrastructures protection[J]. Security and Communication Networks, 2018.

[11] 方滨兴. 从层次角度看网络空间安全技术的覆盖领域[J]. 网络与信息安全学报, 2015, 1(1): 2-7.
FANG B X. A hierarchy model on the research fields of cyberspace security technology[J]. Chinese Journal of Network and Information Security, 2015, 1(1): 2-7.

[12] 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 F, LUO X Y, LIU F L. Research on cyberspace surveying and mapping technology[J]. Chinese Journal of Network and Information Security, 2016, 2(9): 1-11.

[13] 李欲晓, 谢永江. 世界各国网络安全战略分析与启示[J]. 网络与信息安全学报, 2016, 2(1): 1-5.
LI Y X, XIE Y J. Analysis and enlightenment on the cybersecurity strategy of various countries in the world[J]. Chinese Journal of Network and Information Security, 2016, 2(1): 1-5.

[14] 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 L, CAO Y, SU M J, et al. Cyberspace resources surveying and mapping: the concepts and technologies[J]. Journal of Cyber security, 2018, 3(4): 1-14.

[15] 王松, 张野, 吴亚东. 网络拓扑结构可视化方法研究与发展[J]. 网络与信息安全学报, 2018, 4(2): 1-17.
WANG S, ZHANG Y, WU Y D. Survey on network topology visualization[J]. Chinese Journal of Network and Information Security, 2018, 4(2): 1-17.

[16] GOVINDAN R, TANGMUNARUNKIT H. Heuristics for internet map discovery[C]//Proceedings IEEE International Conference on Computer Communications. 2000: 1371-1380.
[17] KEYS K. Iffinder, a tool for mapping interfaces to routers[EB].
[18] SPRING N, MAHAJAN R, WETHERALL D. Measuring ISP topologies with rocketfuel[J]. ACM Sigcomm Computer Communication Review, 2002, 32(4): 133-145.
[19] BENDER A, SHERWOOD R, SPRING N. Fixing ally's growing pains with velocity modeling[C]//Proceedings of the 8th ACM Sigcomm Conference on Internet Measurement. 2008: 337-342.
[20] KEYS K, HYUN Y, LUCKIE M, et al. Internet-scale IPv4 alias resolution with MIDAR[J]. IEEE/ACM Transactions on Networking, 2013, 21(2): 383-399.
[21] SHERWOOD R, SPRING N. Touring the internet in a TCP sidecar[C]//Proceedings of the 6th ACM Sigcomm Conference on Internet Measurement. 2006: 339-344.
[22] SHERRY J, KATZ-BASSETT E, PIMENOVA M, et al. Resolving IP aliases with prespecified timestamps[C]//Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement. 2010: 172-178.
[23] MARCHETTA P, PERSICO V, PESCAPÈ A. Pythia: yet another active probing technique for alias resolution[C]//Proceedings of the 9th ACM Conference on Emerging Networking Experiments and Technologies. 2013: 229-234.
[24] GRAILET J F, DONNET B. Towards a renewed alias resolution with space search reduction and IP fingerprinting[C]//Network Traffic Measurement and Analysis Conference. 2017: 1-9.
[25] GUNES M H, SARAC K. Analytical IP alias resolution[C]//IEEE International Conference on Communications. 2006: 459-464.
[26] AUGUSTIN B, CUVELLIER X, ORGOGOZO B, et al. Avoiding traceroute anomalies with paris traceroute[C]//Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement. 2006: 153-158.
[27] SPRING N, DONTCHEVA M, RODRIG M, et al. How to resolve IP aliases[D]. Seattle: University of Washington, 2004.
[28] 赵洪华, 白华利, 陈鸣, 等. 别名解析中的别名过滤技术[J]. 软件学报, 2009 (8): 2280-2288.
ZHAO H, BAI H L, CHEN M, et al. Alias filtering technique in alias resolution[J]. Journal of Software, 2009 (8): 2280-2288.
[29] TOZAL M, SARAC K. TraceNET: an internet topology data collector[C]//Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement. 2010: 356-368.
[30] PADMANABHAN V N, SUBRAMANIAN L. An investigation of geographic mapping techniques for internet hosts[J]. ACM SIGCOMM Computer Communication Review, 2001, 31(4): 173-185.
[31] GUEYE B, ZIVIANI A, CROVELLA M, et al. Constraint-based geolocation of internet hosts[J]. IEEE/ACM Transactions on Networking, 2006, 14(6): 1219-1232.
[32] SCHAPIRA M, ZHU Y, REXFORD J. Putting BGP on the right path: a case for next-hop routing[C]// Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks. 2010: 3.
[33] LENCSE G, RÉPÁS S. Performance analysis and comparison of different DNS64 implementations for linux, openBSD and freeBSD[C]//IEEE 27th International Conference on Advanced Information Networking and Applications. 2013: 877-884.
[34] ZHAO F, LUO X, GAN Y, et al. IP geolocation based on identification routers and local delay distribution similarity[J]. Concurrency and Computation: Practice and Experience, 2018: 1-15.

[作者简介]

袁福祥（1991– ），男，山东济宁人，信息工程大学博士生，主要研究方向为网络空间资源测绘与 IP 定位。

刘粉林（1964– ），男，江苏溧阳人，博士，信息工程大学教授、博士生导师，主要研究方向为网络空间安全。

刘翀（1994– ），男，辽宁抚顺人，信息工程大学硕士生，主要研究方向为网络空间资源测绘与 IP 定位。

刘琰（1979– ），女，山东济南人，博士，信息工程大学副教授，主要研究方向为网络空间安全。

罗向阳（1978– ），男，湖北荆门人，博士，信息工程大学教授、博士生导师，主要研究方向为网络空间安全。