

国家网络与运营商网络的一致性分析

朱金玉, 张宇, 王宇楠, 张宏莉, 方滨兴

哈尔滨工业大学网络空间安全学院 哈尔滨 中国 150001

摘要 在网络空间主权背景下,以国家主权划分的国家网络与以网络所有权划分的运营商网络之间存在不一致情况。为研究全球各国家的国家网络和运营商网络中互联网基础设施的部署情况,本文从网络拓扑测量数据的角度,提出国家网络和运营商网络的一致性分析方法。首先采用基于路由器约束的启发式方法标注拓扑数据所在的国家与运营商,提取一份全球路由器的国家级“管辖权-所有权”的数据集。其次,提出基于相似率和境外占比差的国家网络与运营商网络的一致性指标,从网络和边界、路由器和连接多维度进行量化,获得一份全球各国家的国家网络与运营商网络的一致性值。最后,从路由器的部署规模、范围和一致性角度研究不同国家互联网基础设施的部署现状和影响。实验结果表明,基于路由器约束的启发式方法标注拓扑数据的覆盖率为99.12%,经互联网交换中心数据验证标注地理国家的准确率为96.47%,运营商国家的准确率为75.08%。在拓扑数据覆盖的全球187个主权国家中,国家网络和运营商网络的境内相似率均值为0.66,有106个国家高于平均值,境外占比差均值为0.15,有121个国家低于平均值。绝大多数国家在网络中路由器和连接的一致性高于在边界一致性。95个国家一致性较高。少数国家拥有着全球互联网大部分的资源,与众多国家的互联网基础设施关联密切,如美国拥有全球路由器总数的39.2%,这些路由器部署在111个国家,而109个国家的路由器部署在5个国家以内。中国、以色列、美国、英国和挪威等国家的互联网基础设施分布在他国的数目远多于他国部署在本国领土的数目,而日本、墨西哥、芬兰、德国等国家恰好相反。

关键词 国家网络; 国家运营商网络; 路由器标注; 一致性; 相似率; 境外占比差

中图分类号 TP393.4 DOI号 10.19363/J.cnki.cn10-1380/tn.2025.03.05

Similarity Analysis of National Network and Operator's Network

ZHU Jinyu, ZHANG Yu, WANG Yunan, ZHANG Hongli, FANG Binxiang

School of Cyberspace Science, Harbin 150001, China

Abstract In the context of cyberspace sovereignty, there is an inconsistency between the National Network divided by national sovereignty and the national Operator's Network divided by network ownership. To study the deployment of the National Network and Operators' Network in various countries around the world, this paper proposes a consistency analysis method for the National Network and Operator's Network from the perspective of network topology measurement data. First, a heuristic method based on router constraints is used to annotate the nation and operator where the topology data is located and a national-level "jurisdiction-ownership" data set of global routers is extracted. Secondly, the consistency index between the National Network and Operator's Network based on the similarity rate and the external ratio difference is proposed, and quantitative analysis is carried out from the multi-dimensional network and border, router and link, and a consistency index between the National Network and the Operator's Network of each nation in the world is obtained. Finally, the deployment status of Internet infrastructure in different nations is studied in terms of the deployment scale, scope, and consistency. The experimental results show that the coverage of the router constraint-based heuristic method for annotating topological data is 99.12%, and the accuracy of annotated geographic countries and operator's countries is 96.47% and 75.08%, respectively, as verified by IXP. Among the 187 sovereign countries in the world covered by the topology data, the average internal similarity between the National Network and Operator's Network is 0.66, with 106 countries above the average, and the average external ratio difference is 0.15, with 121 countries below the average. The vast majority of countries have higher consistency in routers and links in the network than in border routers and links. 95 countries have higher consistency. A few nations have most of the resources of the global Internet and are closely associated with the Internet infrastructure of many nations, for example, the United States has 39.2% of the total number of global routers, which are deployed in 111 nations, while 109 nations have routers deployed in only 5 nations or less. The number of internet infrastructures distributed in other countries such as China, Israel, the United States, the United Kingdom, and Norway far exceeds the number deployed in their own territories, while nations such as Japan, Mexico, Finland, and Germany do the opposite.

通讯作者: 张宇, 博士, 副教授, Email: yuzhang@hit.edu.cn。

本课题得到国家重点研发计划资助项目(No. 22022YFB3102903)和中央高校基本科研业务费专项资金(No. HIT.OCEF.2021007)资助。

收稿日期: 2023-01-17; 修改日期: 2023-08-08; 定稿日期: 2025-01-22

Key words national network; operator's network; routers' annotation; consistency; similarity; external ratio difference

1 引言

网络空间成为继陆地、海洋、天空、外空之外的第五空间。近年来, 各国家在网络空间冲突频发, 2019 年 5 月俄罗斯发布《主权互联网法》并在年底进行国家断网测试^[1]。2020 年 8 月美国发布“清洁网络计划”^[2]。2022 年 2 月美国 Viasat 运营商的 KA-SAT 受 DDOS 攻击, 导致乌克兰多个公共机构和基础设施通信中断, 同时影响了德国、法国意大利等多个国家^[3]。2022 年 3 月 4 日《华盛顿邮报》报道称, 美国 Cogent 运营商停止为俄罗斯客户提供服务^[4]。2022 年 3 月 11 日 LINX 暂停了两家俄罗斯网络的会员资格 Megafon (AS 31133) 和 Rostelecom (AS 12389)^[5]。与此同时, 出于政治和国家安全的考虑, 部分政府强制关闭互联网。从 2022 年初至 8 月全球发生了 14 起网络中断的事件^[6], 例如印度的焦特布尔为了维持法律和秩序断网, 苏丹为解决公立学校考试作弊多次断网, 古巴由于多个城市发生抗议而断网等。

网络已经从促进信息自由流动的无边界网络演变成反映政治和经济约束的网络。在网络空间治理中, 国家网络是以国家地理疆域主权为界, 部署在领土上所有网络实体的集合; 国家运营商网络是以网络实体的所有权为界, 归属于该国家的运营商的所有网络实体的集合。国家网络与运营商网络之间存在不一致的情况, 如图 1 所示。方框表示国家网络 (National Network, NN), 圆形表示国家运营商网络 (Operator's Network, ON), 重合部分表示内部网络 (Internal Network, IN), 不一致部分包括部署在本国的他国境外网络 (Foreign External Network, FE), 其网络实体部署在本国领土上但所有权归属于其他国家; 本国的运营商境外网络 (Operator's External Network, OE), 其网络实体分布在其他国家领土上但所有权归属于本国家。

部署在本国的他国境外网络和分布在他国领土上的本国境外网络中网络实体的管辖权和所有权不一致, 使得网络空间治理方案更加困难。本文首次提出了基于网络拓扑测量数据的国家网络与运营商网络的一致性分析方法, 从量化研究的角度分析每个国家的网络实体的分布情况, 每个国家境内网络与境外网络的边界连通情况, 比较不同国家间网络部署差异。当不同国家间比较国家网络与运营商网络的一致性时, 一致性值越高, 则该国家的网络实体

分布越聚集在本国领土, 管辖权和所有权相同的网络实体占比越高, 此时政府对网络可控性相对较高, 如印度、苏丹和古巴等国家的断网事件。当一致性值越低时, 部署在本国的他国网络实体或分布在他国领土上的本国网络实体较多, 由于某些原因这些网络实体断开与本国或他国的连接, 造成本国或他国网络通信中断。

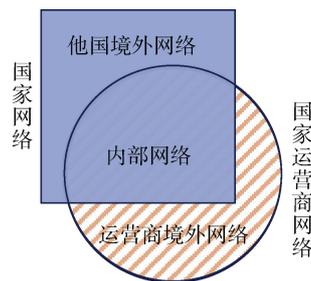


图 1 国家网络与运营商网络示意图

Figure 1 Schematic diagram of national network and operator's network

在以往工作中, 1. 网络拓扑测量作为获得全球各国网络拓扑数据的基础, Motamedi 等人^[7]概述了网络拓扑结构, 包括 IP 级、路由器级、POP 级和 AS 级等; 张宏莉等人^[8]概述了互联网拓扑测量与分析方法; 张宇等人^[9-10]探测了中国 IP 级网络拓扑, 并概述了拓扑建模技术。本文以网络拓扑测量数据为基础, 构建全球互联网基础设施的集合。2. 拓扑结构的地理映射作为划分国家网络的基础数据, Luckie 等人^[11-12]提出从主机名中提取自治域号和地理位置信息的方法; Scheitle 等人^[13]提出采用多个测量框架定位地理信息的方法; 朱金玉等人^[14-15]提出基于多源数据定位路由器地理位置的方法, 依据 Traceroute 路径中直连的路由器位置差异提取国家网络边界; Giotsas 等人^[16]提出映射 IXP 对等连接到机房实体位置的方法。本文以地理定位方法为基础, 提出基于路由器约束的启发式算法对网络拓扑数据定位地理位置, 以国家地理疆域提取国家网络中的互联网基础设施。3. 拓扑结构的 AS 映射作为划分运营商网络的基础数据, Ziv 等人^[17]对 AS 的所有者分类, Carisimo 等人^[18]提出识别 AS 归属的运营商的方法。本文以 AS 映射方法为基础, 提出基于路由器约束的启发式算法对网络拓扑数据映射 AS 的运营商归属信息, 以运营商归属国家提取运营商网络中的互联网基础设施。本文研究专注于分析地理位置上的实体设备构成的网络与 AS 网络拓扑结构的异同, 提出国家网络

和运营商网络的一致性分析方法。

国家网络与运营商网络的一致性分析主要有以下挑战: 1. 难以准确完整地全球网络拓扑数据提取各个国家网络与运营商网络, 网络拓扑数据标注地理位置和运营商国家信息的挑战和基于接口 IP 地址信息标注路由器信息挑战等; 2. 难以从多维度提出国家网络与运营商网络的一致性分析的指标: 从何种角度以何种具体合理的指标量化的一致性的挑战等。

为了解决这些挑战, 本文提出了基于网络拓扑测量数据的多维度的国家网络与国家运营商网络的一致性分析方法。第一阶段以完整准确的标注路由器的地理部署国家和运营商归属国家为目标, 提取一份全球互联网基础设施的国家级“管辖权-所有权”的数据集; 第二阶段从多维度量化国家网络与运营商网络的一致性为目标, 从网络和边界两个角度, 从相似率和境外占比差两个方面提出一致性指标, 量化全球各个国家的国家网络与运营商网络的一致性值; 第三阶段以分析全球互联网基础设施的部署现状为目标, 从部署规模、范围、一致性角度分析全球各个国家互联网基础设施的部署情况。

本文的创新点在于: 1. 以网络拓扑数据为切入点, 将路由器数目类比互联网基础设施数目, 探讨网络空间主权中各个国家的国家网络和运营商网络中互联网基础设施的一致性, 基于互联网基础设施的物理位置和所有权归属的异同分析全球互联网基础设施的部署现状; 2. 标注方法的创新点在于: 提出基于路由约束的启发式方法标注公开数据集的路由器的地理位置和运营商信息, 易于实施与重复, 结果易于更新与评估; 3. 分析方法的创新点在于: 从网络的角度分析各国网络的部署现状, 从边界的角度分析各国网络间的互联互通, 以相似率作为国家和运营商网络的境内一致率指标, 以境外占比差作为国家网络和运营商网络的境外差异性指标, 量化各国的国家与运营商网络的一致性值, 易于比较不同国家的互联网基础设施部署的异同。

文章结构如下: 第二节为研究框架, 构建国家网络与运营商网络, 提出分析国家网络与运营商网络的一致性指标的关系模型; 第三节为网络拓扑标注, 为拓扑数据标注地理国家和运营商国家; 第四节为一致性分析, 提出国家网络与运营商网络的相似率和境外占比差指标; 第五节为互联网基础设施部署现状分析, 提出以路由器作为互联网基础设施分析各个国家的部署差异影响; 第六节总结全文。

2 研究框架

2.1 模型与问题

全球网络空间的拓扑数据表示为 $G=(V,L)$, G 表示网络拓扑图, V 表示图中的点的集合, 即所有路由器的集合 $V=\{R_i|i \in \mathbb{N}\}$, L 表示 G 中边的集合, 即路由器间连接的集合 $L=\{(R_i,R_j)|R_i,R_j \in V\}$ 。标注路由器的地理位置 $N(R)$, 标注路由器的运营商国家 $O(R)$, 提取国家 K 的网络 $G_N=(R_N,L_N)$, 提取国家 K 的运营商网络 $G_O=(R_O,L_O)$ 。内部网络 $G_I=G_N \cap G_O$, 部署在本国领土的他国境外网络 $G_{FE}=G_N - G_I$, 本国的运营商境外网络 $G_{OE}=G_O - G_I$ 。

在全球网络拓扑结构中, 边界网络是不同网络间互联互通的关键基础设施^[11]。当以国家地理疆域主权为边界划分的国家网络间存在国家网络的边界网络 G_{Nbd} , 与他国网络连接的边界连接 $L_{Nbd}=\{(R_i,R_j)|R_i \in G_N,R_j \notin G_N\}$, 提取国家网络边界路由器 R_{Nbd} 。当以网络实体的所有权为边界划分的国家运营商网络间存在国家运营商网络的边界网络 G_{Obdr} , 与他国运营商网络相连的边界连接 $L_{Obdr}=\{(R_i,R_j)|R_i \in G_O,R_j \notin G_O\}$, 提取国家运营商网络的边界路由器 R_{Obdr} 。

在分析国家网络与运营商网络的一致性时, 从网络角度分析国家网络和运营商网络的部署规模, 从边界角度分析与他国的互联互通, 以相似率量化境内一致性, 以境外占比差量化境外差异性。相似率是国家网络与运营商网络中路由器和连接或边界路由器和连接相同的数目与总数目的比例, 表示境内部署数目的一致性。境外占比差是国家网络或边界网络中他国境外网络占比与运营商网络或边界网络中本国运营商境外网络占比的差值, 表示境外部署数目的差异性。 G_I 越大, G_{FE} 和 G_{OE} 越小, 相似率越大, G_N 与 G_O 境内部署数目越一致; G_{FE} 占比 G_N 与 G_{OE} 占比 G_O 越相似时, 境外占比差越小, G_N 与 G_O 境外部署数目越一致。国家网络与运营商网络在网络和边界角度的相似率越大, 一致性越大, 当相似率相近时, 境外占比差越小, 一致性越大。

在研究全球互联网基础设施部署情况时, 国家网络与运营商网络的相似率越大, 本国运营商的基础设施部署在本国的越多越集中; 境外占比差越小, 部署在他国和他国网络分布在本国的基础设施数目越相近越均衡。

2.2 方案概览

分析国家网络与运营商网络的一致性分为三个阶段, 如图2所示:

阶段一: 标注网络拓扑的地理位置和运营商国家。针对难以准确完整地将全球网络拓扑数据划分为国家网络与运营商网络的问题, 提出基于路由器

约束的启发式算法, 采用公开数据集如: 路由器级拓扑数据、地理定位库、BGP数据、AS2Country数据标注路由器接口IP地址, 以路由器接口IP地址和连接的约束标注路由器的地理国家和运营商国家, 提取全球各个国家的国家网络和运营商网络的路由器和连接, 详见3节。

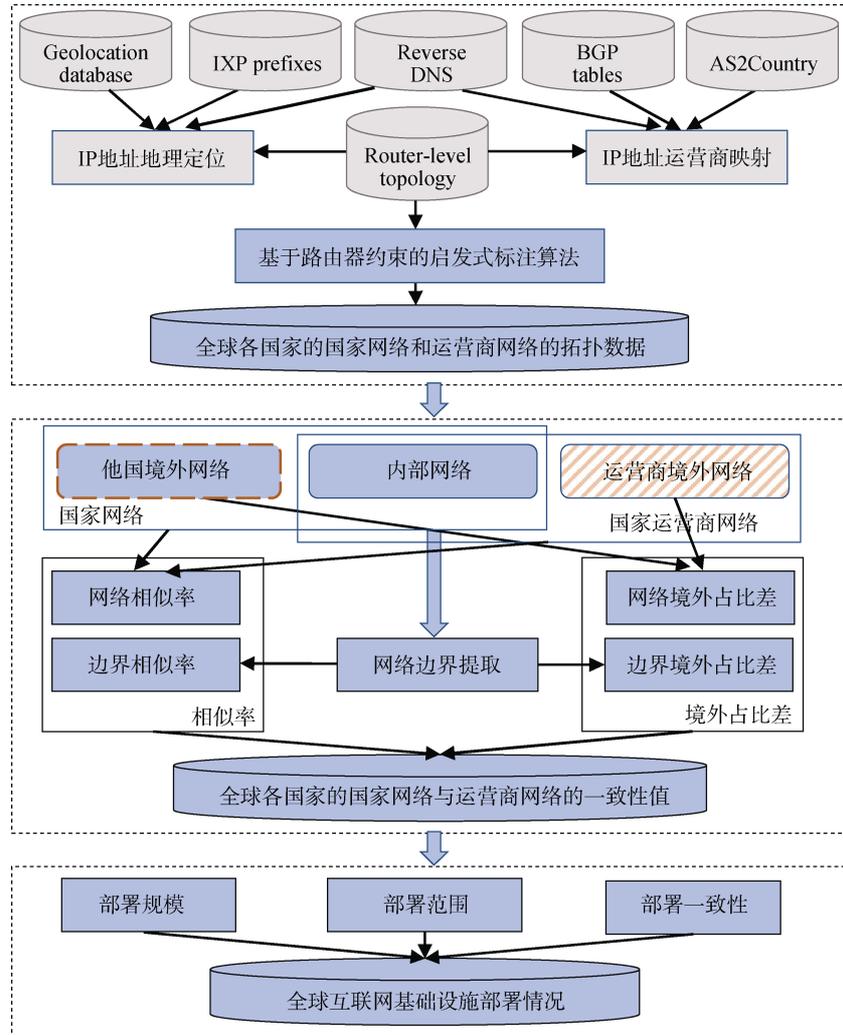


图2 国家网络与运营商网络一致性分析概览

Figure 2 Overview of national network and operators' network consistency analysis

阶段二: 量化国家网络与运营商网络的一致性。针对难以从多维度提出国家网络与运营商网络的一致性指标的问题, 从网络和边界两个角度分析国家网络和运营商网络的一致性, 从相似率和境外占比差两个部分提出一致性指标, 从路由器和连接两个方面量化指标, 获得全球各个国家的国家网络与运营商网络的一致性值, 详见4节。

阶段三: 分析全球互联网基础设施的部署现状。针对难以评估不同国家网络基础设施的部署情况, 以路由器比作互联网基础设施, 从路由器部署规模、范围、一致性角度分析各个国家互联网基础设施部

署情况, 以一致性平均值为界, 将全球各个国家分为4类比较不同国家间部署差异, 详见5节。

3 网络拓扑标注

在路由器级网络拓扑 $G=(V,L)$ 中, V 中路由器 R 有1个或多个接口IP地址 $R_i = \{IP_1, IP_2, \dots, IP_n\}$, 对接口IP地址标注地理位置信息 $N(IP_i)$ 和运营商国家信息 $O(IP_i)$, 采用基于路由器约束的启发式方法 $C(R)$ 标注路由器的地理位置 $C_N(R)$ 和运营商国家 $C_O(R)$, 可标注的路由器为 R' , 已知地理位置或运营商国家的

路由器为地标 R_{gr} 。标注的目标是在保证准确率的前提下, 提高标注方法的覆盖率。

网络拓扑标注国家信息主要步骤如下 §3.1 路由器级拓扑数据清洗; §3.2 路由器标注方法; §3.3 路由器标注结果; §3.4 局限性。

3.1 路由器级拓扑数据清洗

CAIDA ITDK 每年公开 1-2 个版本的全球路由器级网络拓扑数据。在 Ark 测量平台对 IPv4 的/24 BGP 前缀中随机选择目的 IP 地址进行 tTraceroute 测量得到 IP 级拓扑, 采用 MIDAR 和 Iffinder 等别名解析技术生成路由器级拓扑^[19]。本文以研究不同国家互联网基础设施的部署现状为目标, 以路由器为实验数据, 不考虑终端节点。采用 CAIDA ITDK 2020 年 8 月发布的路由器级拓扑数据, 该数据中包含一些非真实的地址, 在标注前将其清洗。

清洗数据的规则有三个: 1. 删除终端节点: 在 tTraceroute 拓扑测量中, 如节点只作为目的跳出现而未在中间路径中出现时, 将其视为终端节点; 2. 删除私有 IP 地址: CAIDA 数据将探测到的私有 IP 地址转化为 224.0.0.0/3 网段的 IP 地址; 3. 删除运营商级 NAT 地址: 数据中存在 100.64.0.0/16 网段的运营商级 NAT 地址无法标记地理位置及运营商国家。

3.2 路由器标注方法

针对难以准确完整的标注路由器级网络拓扑的地理位置和运营商国家的问题, 本文提出了基于路由器约束的启发式方法将 IP 级网络拓扑的地理位置和运营商国家映射到路由器。依据事实, 我们提出启发式规则: 1. 路由器有多个接口 IP 地址但地理位置和运营商国家只有一个; 2. 邻居路由器间地理位置和运营商国家相近。

基于路由器约束的启发式方法包括 3 个步骤:

步骤 1. 接口选择, 通过标注路由器接口 IP 地址推断路由器的地理位置和运营商国家。在标注路由器接口 IP 地址时可能出现 4 种情况, 1. 路由器可直接标注, 其接口 IP 地址标注信息全部相同; 2. 路由器可直接推断标注, 其多个接口 IP 地址标注的地理位置或运营商国家不相同, 但大多数标注结果相同; 3. 路由器无法标注, 其多个接口 IP 地址标注的地理位置或运营商国家不相同; 4. 路由器的接口 IP 地址无法标注, 其接口 IP 地址在地理定位库或 IP2AS 库中无法定位。

基于路由器有多个接口 IP 地址但地理位置和运营商国家只有一个的事实, 将每个接口 IP 地址标注的国家权重设为 1, (对于待标注的路由器而言, 其每

个接口 IP 地址的权重均相同)选择频数最高的国家为路由器的标注结果。步骤 1 可将情况 1 和 2 中的路由器标注, 未标注路由器中, 情况 3 为位置冲突, 情况 4 为无位置。

步骤 2. 邻居路由器选择, 通过网络拓扑中已标注的路由器的信息推断未标注的路由器的地理位置和运营国家。在路由器标注时可能出现 4 种情况, 1. 路由器可直接标注, 其邻居路由器的标注信息相同; 2. 路由器可直接推断标注, 其邻居路由器的标注信息不相同, 但大多数相同; 3. 路由器无法标注, 其邻居路由器的标注信息不同; 4. 邻居路由器无信息, 其邻居路由器均在步骤 1 中无法标注。

基于邻居路由器间地理位置和运营商国家应相近的事实, 标注步骤 1 中未标注的路由器。将邻居路由器的标注信息的权重均设为 1, (对于待标注的路由器而言, 其每个邻居路由器的权重均相同)选择频数最高的国家为路由器的标注结果, 不断迭代至没有新的可标注路由器。步骤 2 可将情况 1 和 2 中的路由器标注, 未标注路由器中, 情况 3 为位置冲突, 情况 4 为无位置。

步骤 3. 邻居路由器接口选择, 通过网络拓扑中已标注的路由器及其接口 IP 地址信息推断未标注的路由器的地理位置和运营商国家。在通过邻居路由器接口选择标注时可能出现 3 中情况, 1. 路由器可直接推断标注, 其邻居路由器的所有接口 IP 地址标注信息不相同, 但大多数相同; 2. 路由器无法标注, 其邻居路由器的所有接口 IP 地址标注信息不相同; 3. 邻居路由器无信息, 其邻居路由器均在步骤 1、2 中无法标注。

基于启发式规则 1 和 2, 标注 2 中未标注的路由器。将未标注的路由器的所有邻居路由器的每个接口 IP 地址标注的国家权重设为 1, (对于待标注的路由器而言, 其每个邻居路由器的接口 IP 地址的权重均相同, 但每个邻居路由器的权重不同, 接口多的路由器权重更大)选择频数最高的国家为路由器的标注结果, 不断迭代至没有新的可标注路由器。步骤 3 可将情况 1 中的路由器标注, 未标注路由器中, 情况 2 为位置冲突, 情况 3 为无位置。

基于路由器约束的启发式方法无法标注两类路由器, 路由器的接口 IP 地址和邻居路由器及其接口 IP 地址的无位置信息或位置信息中均无法获得最高频数位置。

算法 1. 基于路由器约束的标注方法

输入: 路由器的集合 R , R_i 的接口 IP 地址集 IP_i , IP 地

理定位库, BGP 路由器表, IP2Country 映射表

输出: 带标注信息的路由器集合 R'

0. 预处理: 标注路由器接口 IP 地址

FOR each R_i in R :

FOR each IP_j in R_i :

$N(IP_j) \rightarrow R_i^N$

$O(IP_j) \rightarrow R_i^O$

1. 接口选择:

FOR each R_i in R :

IF R_i^N 频数最多的国家 K THEN

$C(R_i) = K$

ELSE $R_i \rightarrow \neg R_{interface}$

2. 邻居路由器选择:

FOR each R_i in $\neg R_{interface}$:

R_i 邻居路由器位置为 $R_{i_neighbor}^N$

IF $R_{i_neighbor}^N$ 频数最多的国家 K THEN

$C(R_i) = K$

ELSE $R_i \rightarrow \neg R_{neighbor_router}$

3. 邻居路由器接口选择:

FOR each R_i in $\neg R_{neighbor_router}$:

R_i 邻居路由器接口位置为 $R_{i_neighbor_interface}^N$

IF $R_{i_neighbor_interface}^N$ 频数最多的国家 K THEN

$C(R_i) = K$

ELSE $R_i \rightarrow \neg R_{neighbor_interface}$

4. 重复 1-3 标注路由器的运营商国家

国家^[13]。

采用基于路由器约束的启发式方法标注结果如表 1 所示。在标注路由器运营商国家时, 根据 pPeeringdb 公开数据删除路由器中 IXP 的 IP 地址。IXP 前缀为参与交换通信的 AS 分配 IP 地址, 使基于路由器约束的启发式方法 AS 标注错误。

在标注地理位置中, 步骤 1. 接口选择标注路由器 2,619,481 个, 接口无位置 5,083 个, 接口位置冲突 3,090 个; 步骤 2. 邻居路由器选择标注路由器 7,304 个, 邻居路由器无位置 476 个, 位置冲突 393 个; 步骤 3. 邻居接口选择标注路由器 162 个, 邻居路由器接口无位置 476 个, 位置冲突 231 个。

在标注运营商国家中, 步骤 1. 接口选择标注路由器 2,522,770 个, 接口无位置 91,360 个, 接口位置冲突 13,524 个; 步骤 2. 邻居路由器选标注路由器 78,280 个, 邻居路由器无位置 20,173 个, 位置冲突 6,363 个; 步骤 3. 邻居接口选择标注路由器 3,365 个, 邻居路由器接口无位置 20,172 个, 位置冲突 3,067 个。

3.3.1 覆盖率

覆盖率是指可以标注的路由器数目占需要标注的路由器总数目的比例。需要标注的路由器数目是清洗后的 CAIDA 公开路由器级拓扑数据, 包括单接口路由器和多接口路由器。

基于路由器约束的启发式方法的覆盖率如表 2 所示。在地理位置标注中, 步骤 1 接口选择的覆盖率为 99.69%, 步骤 2 邻居路由器选择的覆盖率为 89.37%, 步骤 3 邻居接口选择的覆盖率为 18.64%, 基于路由器约束的启发式方法整体覆盖率为 99.97%。在运营商国家标注中, 步骤 1 接口选择的覆盖率为 96.01%, 步骤 2 邻居路由器选择的覆盖率为 74.68%, 步骤 3 邻居接口选择的覆盖率为 12.65%, 基于路由器约束的启发式方法总覆盖率为 99.12%。

3.3.2 准确率

准确率是指已标注的路由器与地标点路由器标注结果相同的数目占地标点路由器的比例。公开的 IXP 前缀库^[23]可以映射部分 IP 地址的地理位置

3.3 路由器标注结果

对于大规模网络的 IP 地址的地理定位和运营商国家映射而言, 综合时间和效率等问题, 被动的标注方法较主动测量的方法更适用^[15]。为标注清洗后的路由器的每个接口 IP 地址, 采用 Maxmind 免费的 IP 地理定位库标注地理位置^[20], 采用 RouterViews 和 RIPE 的 RIS 在 2020 年 8 月的 BGP 前缀数据, 将接口 IP 地址映射 IP2AS^[21]和 AS2Country^[22]。AS 被单一运营商管理^[12], AS 归属的国家也是运营商的

表 1 路由器地理国家标注和运营商国家标注

Table 1 Annotation of router's territory nation and operators' nation

	地理国家标注			运营商国家标注		
	接口选择	邻居路由器选择	邻居路由器接口选择	接口选择	邻居路由器选择	邻居路由器接口选择
可标注路由器	2,619,481	7,304	162	2,522,770	78,280	3,365
未标注	无位置	476	476	91,360	20,173	20,172
	位置冲突	3,090	393	231	13,524	6,363

表 2 基于路由器约束的启发式标注方法覆盖率

Table 2 The coverage of heuristic method to annotate routers based on router constraint

	接口选择	邻居路由器选择	邻居路由器接口选择	R-constraint
国家	99.69%	89.37%	18.64%	99.97%
运营商国家	96.01%	74.68%	12.65%	99.12%

和 AS 信息^[14], 反向域名数据(reverse DNS, rDNS)可提取部分 IP 地址隐含的地理位置^[16-17]和 AS 信息^[18]。从 IXP 和 rDNS 中提取的 IP 地址的国家地理位置和运营商国家信息可作为地标数据验证标注的准确率。

选择 CAIDA 公开的 IXP 数据中所有机房均在同

一国家的路由器作为 IXP 的地标点, 共获取 19,581 个运营商网络的地标点和 19,560 个国家网络的地标点。采用 CAIDA 的 HOIHO 方法选择可提取 rDNS 的正则表达式和时延信息的路由器作为 rDNS 地标点, 共获取 4,767 个运营商网络的地标点和 54,209 个国家网络的地标点。

以地标点为真值数据对基于路由器约束的启发式方法验证每一步骤的准确率。为评价原始数据的局限性, 本文提出了真值极大 Max 和真值级小 Min 的比较方法。Max 是任意路由器接口 IP 地址标注信息与地标点相同则标注正确; Min 是任意路由器接口 IP 地址标注信息与地标点不同则标注错误, 各方法准确率如表 3 所示。

表 3 基于路由器约束的启发式方法标注路由器的准确率

Table 3 The accuracy of heuristic method to annotate routers based on router constraint

			接口选择	邻居路由器选择	邻居路由器接口选择	R-constraint
国家	rDNS	准确率	75.47%	38.50%	74.92%	75.37%
		Max	76.00%	81.00%	76.06%	76.01%
		Min	74.74%	3.21%	73.69%	74.54%
	IXP	准确率	96.62%	41.46%	96.27%	96.47%
		Max	97.35%	95.14%	97.34%	97.34%
		Min	95.84%	0	95.24%	95.57%
运营商国家	rDNS	准确率	43.12%	43.21%	43.18%	43.12%
		Max	43.88%	63.28%	56.02%	44.48%
		Min	32.87%	1.22%	10.88%	31.89%
	IXP	准确率	76.00%	46.72%	46.89%	75.08%
		Max	78.71%	68.98%	67.68%	78.40%
		Min	32.43%	0.61%	1.8%	31.43%

本文在仅使用公开免费有限数据且未使用公开的地标数据和主动测量等数据的前提下, 基于路由器约束的启发式方法在标注国家地理位置的准确率均达到原始标注数据极大值的 99%以上(即为准确率/真值极大 Max 准确率, rDNS 和 IXP 验证分别为 75.37%/76.01%和 96.47%/97.34%); 在标注运营商国家的准确率达到原始标注数据极大值的 96%以上(即为准确率/真值极大 Max 准确率, rDNS 和 IXP 验证分别为 43.12%/44.48%和 75.08%/78.40%), 验证了方法的准确性。但本方法的准确性仍依赖于公开数据的准确性, IXP 验证准确率在国家地理位置的准确率为 96.47%, 在运营商国家的准确率为 75.08%。rDNS 验证准确率较低, 在国家地理位置的准确率为 75.37%, 在运营商国家的准确率仅为 43.12%。

3.4 局限性

本文提出基于路由器约束的启发式方法, 采用 2020 年 8 月的 Maxmind 免费定位库和 RouterViews

和 RIPE 的 RIS 的 BGP 前缀数据标注 CAIDA ITDK 同时期的路由器级拓扑数据的地理位置和运营商国家, 提取了一份全球路由器的国家级“管辖权-所有权”的数据集, 共计 2,604,400 个路由器。基于路由器约束的启发式方法标注地理位置和运营商归属的覆盖率达 99.97%和 99.12%, 准确率达到原始标注数据准确率极大值的 99%以上和 96%以上。

基于路由器约束的启发式方法的标注结果一定程度上受 IP 定位库和 IP2AS 的准确率和覆盖率影响。如果路由器的接口 IP 地址和其邻居路由器的接口采用 IP 定位库或 IP2AS 均无法标注, 则基于路由器约束的启发式方法无法对路由器进行标注。只有一个接口 IP 地址的路由器标注结果的准确率和覆盖率完全依赖于公开免费库的质量。

4 一致性分析

从标注地理国家和运营商国家的路由器级网络

拓扑中, 提取每个国家的国家网络与运营商网络。图 3 表示不同网络的路由器和路由器间连接。正方形表示路由器, 路由器间连线表示路由器间连接, 蓝色实线表示本国, 黄色虚线表示他国。

从网络的角度分析国家网络和运营商网络的部署规模。 I_i 是国家和运营商的网络相同的路由器, FE_i 是部署在本国的他国网络的路由器, OE_i 是国家运营商网络中分布在境外网络的路由器。从边界的角度分析国家网络和运营商网络与不同国家互联互通。 $L(I_4, E_4)$ 与 $L(OE_4, FE_4)$ 是 NN 和 ON 的边界相同的连接, $L(I_1, OE_1)$ 与 $L(FE_1, E_1)$ 分别是 OE 与 FE 的国家网络的边界连接, $L(I_2, FE_2)$ 与 $L(OE_2, E_2)$ 是 FE 和 OE 的运营商网络的边界连接。

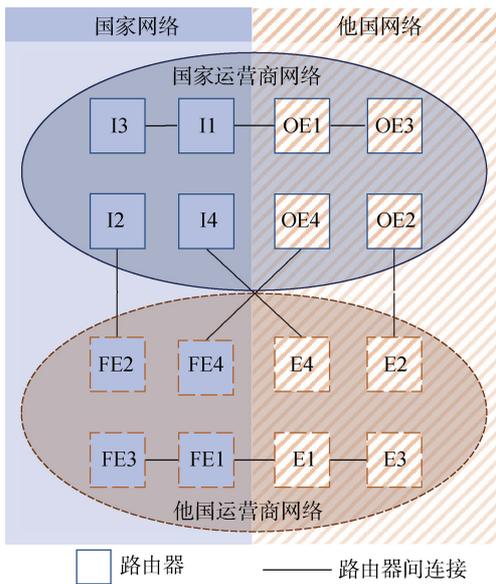


图 3 国家网络与运营商网络中路由器和连接示意图
Figure 3 Schematic diagram of routers and links in national network and operator's network

基于相似率提出国家网络和运营商网络的境内一致性指标, I_i 越多 $L(I_4, E_4)$ 越多, OE_i 越少 $L(I_1, OE_1)$ 与 $L(OE_2, E_2)$ 越少, FE_i 越少 $L(I_2, FE_2)$ 与 $L(FE_1, E_1)$ 越少, 相似率越高。基于境外占比差提出国家网络和运营商网络的境外差异性指标, OE_i 与 FE_i 数目越相近, $L(I_1, OE_1)$ 和 $L(FE_1, E_1)$ 与 $L(I_2, FE_2)$ 和 $L(OE_2, E_2)$ 越相近, 境外占比差越小。从路由器和连接两个方面量化国家网络与运营商网络的一致性值。

本章主要步骤如下 §4.1 网络一致性分析; §4.2 边界一致性分析; §4.3 国家网络和运营商网络的一致性; §4.4 局限性。

4.1 网络一致性分析

从网络的角度在路由器和连接两个方面量化国家与国家运营商的网络一致性指标。网络一致性分

析示意图如图 4 所示, 假设国家网络和运营商网络中路由器和连接数目固定不变, 在图 4-1 中根据 I 中路由器和连接数目差异分 A、B、C 三种情况, 情况 C 的 I 最大, OE 和 FE 最小, 网络相似率最大。在图 4-2 中国家和国家运营商的网络相似率相同, 根据 OE 与 FE 中路由器和连接数目差异分 B1、B2、B3 三种情况。情况 B1 中 OE 占 ON 与 FE 占 NN 比例差异最小且无差异, B1 的境外占比差最小。在 A、B、C 中, C 的相似率最大, 一致率最大; 在情况 B 中, B1、B2、B3 相似率相同, B1 的境外占比差最小, 一致率最大。

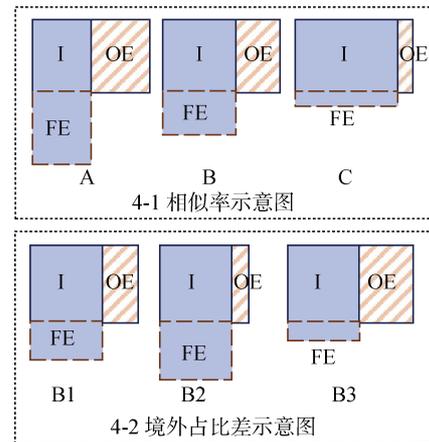


图 4 国家网络与运营商网络的网络一致性分析
Figure 4 Schematic diagram of network consistency analysis between national network and operator's network

极端情况下, 国家网络和运营商网络中无 OE 和 FE, 则相似率为 1, 境外占比差为 0, 网络相似率最大、境外占比差最小、一致性最大。

4.1.1 网络相似率

以 Jaccard 系数量化国家网络和国家运营商网络的网络一致率 J_{NS} 可表示为:

$$J_{NS} = \frac{G_N \cap G_O}{G_N \cup G_O} = \frac{G_I}{G_I + G_{OE} + G_{FE}} \quad (4.1.1)$$

从路由器和连接两个方面分别量化国家网络和运营商网络的一致性 $J_{NS,R}$ 和 $J_{NS,L}$, 以路由器和连接间的关系即节点度关联路由器和连接指标。每个连接将两个路由器关联成邻居路由器, 如果一个路由器 R 与 m 个路由器连接, 则路由器 R 有 m 个连接, 其节点度为 m 。国家网络中路由器的平均节点度为 d_N , 国家运营商网络中路由器的平均节点度为 d_O , 以 d_N 和 d_O 的几何平均数为该国家路由器的平均节点度 $\sqrt{d_N + d_O}$ 。以每个连接的权重为 1, 则路由器的权重为其节点度, 将权重进行归一化处理得到路由器

和连接的一致率系数 γ_R 和 γ_L 。其中, 路由器的节点度越高, 在路由器方面的相似率权重越高, 连接方面的相似率权重越低。

$$J_{NS} = J_{NS_R} * \gamma_R + J_{NS_L} * \gamma_L \quad (4.1.1.1)$$

$$J_{NS_R} = \frac{R_N \cap R_O}{R_N \cup R_O} = \frac{R_I}{R_I + R_{OE} + R_{FE}} \quad (4.1.1.2)$$

$$J_{NS_L} = \frac{L_N \cap L_O}{L_N \cup L_O} \quad (4.1.1.3)$$

路由器的节点度 d_{R_i} 是路由器 R_i 无向边的条数, 国家网络和运营商网络中路由器平均节点度 d_N 和 d_O 分别表示为:

$$d_N = \frac{1}{|R_N|} \sum_{\forall R_i \in R_N} d_{R_i} \quad (4.1.1.4)$$

$$d_O = \frac{1}{|R_O|} \sum_{\forall R_i \in R_O} d_{R_i} \quad (4.1.1.5)$$

路由器和连接两方面量化指标的权重 γ_R 和 γ_L 表示为:

$$\gamma_R = \frac{\sqrt{d_N * d_O}}{1 + \sqrt{d_N * d_O}} \quad (4.1.1.6)$$

$$\gamma_L = \frac{1}{1 + \sqrt{d_N * d_O}} \quad (4.1.1.7)$$

$$\gamma_R + \gamma_L = 1 \quad (4.1.1.8)$$

以图 3 中路由器和连接示意图表示 L_N 和 L_O 为:

$$L_N = \left\{ \begin{array}{l} (R_{I_1}, R_{I_3}), (R_{I_1}, R_{OE_1}), (R_{I_2}, R_{FE_2}), (R_{I_4}, R_{E_4}), \\ (R_{FE_4}, R_{OE_4}), (R_{FE_1}, R_{FE_3}), (R_{FE_1}, R_{E_1}) \end{array} \right\}$$

$$L_O = \left\{ \begin{array}{l} (R_{I_1}, R_{I_3}), (R_{I_1}, R_{OE_1}), (R_{OE_1}, R_{OE_3}), (R_{I_4}, R_{E_4}), \\ (R_{I_2}, R_{FE_2}), (R_{OE_4}, R_{FE_4}), (R_{OE_2}, R_{E_2}) \end{array} \right\}$$

其中, $L_I \neq L_N \cap L_O$, 连接 (R_{FE}, R_{OE}) 既是国家网络的连接也是国家运营商网络的连接, 但不是国家运营商境内网络的连接。

4.1.2 网络境外占比差

国家网络中他国运营商境外网络占比 J_N 和国家运营商网络中国家运营商境外网络占比 J_O , 境外占比差 J_{ND} 表示为:

$$J_{ND} = |J_{NN} - J_{NO}| = \left| \frac{G_{FE}}{G_N} - \frac{G_{OE}}{G_O} \right| \quad (4.1.2)$$

从路由器和连接两个方面分别量化境外网络占国家网络和运营商网络的境外占比差 J_{ND_R} 和 J_{ND_L} , 以节点度关联指标如公式(4.1.1.6)和式(4.1.1.7)。

$$J_{ND} = J_{ND_R} * \gamma_R + J_{ND_L} * \gamma_L \quad (4.1.2.1)$$

$$J_{ND_R} = \left| \frac{R_{FE}}{R_N} - \frac{R_{OE}}{R_O} \right| \quad (4.1.2.2)$$

$$J_{ND_L} = \left| \frac{L_{FE}}{L_N} - \frac{L_{OE}}{L_O} \right| \quad (4.1.2.3)$$

以图 3 中路由器和连接示意图表示 L_{OE} 和 L_{FE} 为:

$$L_{OE} = \left\{ \begin{array}{l} (R_{I_1}, R_{OE_1}), (R_{OE_1}, R_{OE_3}), \\ (R_{OE_4}, R_{FE_4}), (R_{OE_2}, R_{E_2}) \end{array} \right\}$$

$$L_{FE} = \left\{ \begin{array}{l} (R_{I_2}, R_{FE_2}), (R_{FE_4}, R_{OE_4}), \\ (R_{FE_1}, R_{FE_3}), (R_{FE_1}, R_{E_1}) \end{array} \right\}$$

4.2 边界一致性分析

从边界的角度在路由器和连接两个方面量化国家网络和运营商网络的边界一致性。边界一致性分析示意图如图 5 所示, 国家运营商境内网络的边界网络 I_{bdr} , 他国运营商境外网络的边界网络 FE_{bdr} , 本国运营商的境外网络 OE_{bdr} 。国家网络的边界网络 N_{bdr} 是 I_{bdr} 和 FE_{bdr} 的国家网络边界, 不包括 $L(I_{bdr}, FE_{bdr})$; 运营商网络的边界网络 O_{bdr} 是 I_{bdr} 和 OE_{bdr} 的运营商网络边界, 不包括 $L(I_{bdr}, OE_{bdr})$ 。当 I_{bdr} 越多, OE_{bdr} 和 FE_{bdr} 的越少时, 边界相似率越大, OE_{bdr} 和 FE_{bdr} 数目越相近时, 边界境外占比差越小, 一致性越大。

在极端情况下, FE_{bdr} 和 OE_{bdr} 无边界, 均为 I_{bdr} 和 E_{bdr} 间的边界, 相似率为 1, 境外占比差为 0, 边界相似率最大、境外占比差最小、一致性最小。

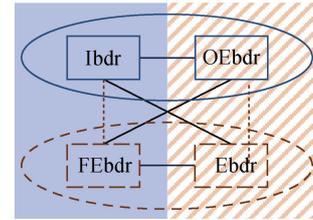


图 5 国家网络与运营商网络的边界一致性分析图
Figure 5 Schematic diagram of border consistency analysis between national network and operator's network

4.2.1 边界相似率

以 Jaccard 系数量化国家和国家运营商的边界相似率 J_{bdrS} 可表示为:

$$\begin{aligned} J_{bdrS} &= \frac{G_{Nbdr} \cap G_{Obdr}}{G_{Nbdr} \cup G_{Obdr}} \\ &= \frac{(G_{Ibdr} \cup G_{FEbdr}) \cap (G_{Ibdr} \cup G_{OEbdr})}{G_{Ibdr} \cup G_{FEbdr} \cup G_{OEbdr}} \end{aligned} \quad (4.2.1)$$

从路由器和连接两个方面分别量化国家网络和

运营商网络的边界相似率 J_{bdrS_R} 和 J_{bdrS_L} , 以节点度关联指标如公式(4.1.1.1)。

以图 3 中路由器和连接示意图表示 L_{Nbdr} 和 L_{Obdr} 为:

$$L_{Nbdr} = \left\{ \begin{array}{l} (R_{I_1}, R_{OE_1}), (R_{I_4}, R_{E_4}), \\ (R_{OE_4}, R_{FE_4}), (R_{FE_1}, R_{E_1}) \end{array} \right\}$$

$$L_{Obdr} = \left\{ \begin{array}{l} (R_{I_2}, R_{FE_2}), (R_{I_4}, R_{E_4}), \\ (R_{OE_4}, R_{FE_4}), (R_{OE_2}, R_{E_2}) \end{array} \right\}$$

R_{Nbdr} 是 L_{Nbdr} 中的 R_{I_1} 、 R_{I_4} 、 R_{FE_4} 和 R_{FE_1} 的集合;
 R_{Obdr} 是 L_{Obdr} 中的 R_{I_2} 、 R_{I_4} 、 R_{OE_4} 和 R_{OE_2} 的集合。

4.2.2 边界境外占比差

国家网络的边界中仅为国家网络的边界占比 J_{Nbdr} 和国家运营商网络的边界中仅为国家运营商网络的边界占比 J_{Obdr} , 比较边界境外占比差 J_{bdrD} 表示为:

$$J_{bdrD} = |J_{Nbdr} - J_{Obdr}|$$

$$= \left| \frac{G_{Nbdr} - G_{Nbdr} \cap G_{Obdr}}{G_{Nbdr}} - \frac{G_{Obdr} - G_{Nbdr} \cap G_{Obdr}}{G_{Obdr}} \right| \quad (4.2.2)$$

$$= \left| \frac{G_{Nbdr} \cap G_{Obdr}}{G_{Nbdr}} - \frac{G_{Nbdr} \cap G_{Obdr}}{G_{Obdr}} \right|$$

从路由器和连接两个方面分别量化国家网络和运营商网络的境外占比差 J_{ND_R} 和 J_{ND_L} , 以节点度关联指标如公式(4.1.2.1)。

4.3 一致性

采用 F-Measure 调和平均数综合评价从网络和边界的角度国家网络与运营商网络的一致性, 国家网络和运营商网络的相似率 J_S 和境外占比差 J_D 可表示为:

$$J_S = \frac{(\omega^2 + 1) * J_{NS} * J_{bdrS}}{\omega^2 * J_{NS} + J_{bdrS}} \quad (4.3.1)$$

$$J_D = \frac{(\omega^2 + 1) * J_{ND} * J_{bdrD}}{\omega^2 * J_{ND} + J_{bdrD}} \quad (4.3.2)$$

其中, $\omega=1$ 时, 网络和边界两个角度量化指标的权重相同, $\omega>1$ 时, 边界角度量化指标的权重大于网络角度。本文设 $\omega=1$, 网络和边界两个角度量化指标权重相同, 国家的 J_{NS} 和 J_{bdrS} 越高, J_S 越高; 国家的 J_{ND} 和 J_{bdrD} 越低, J_D 越低。任意国家的 J_S 越高, 国家网络与运营商网络越一致; 当不同国家 J_S 相近时, J_D 越低越一致。

4.3.1 相似率结果

国家网络与运营商网络的相似率结果如图 6 所示, 图 6 和图 7 分别表示相似率频数分布和各国家对应的相似率(图中以国家网络与运营商网络的相似率降序

排列, 横坐标国家为每七个国家中的第一个), 国家网络与运营商网络的相似率平均值为 **0.66**, 107 个国家超过平均值, 在 0.9-1 的区间分布最多为 34 个国家; 网络相似率平均值为 **0.82**, 123 个国家超过平均值, 在 0.9-1 的区间分布最多为 93 个国家; 边界相似率平均值为 **0.58**, 96 个国家超过平均值, 在 0.7-0.8 区间分布最多为 26 个国家。图 7 展示了每个国家在网络和边界的相似率和国家网络和运营商网络的相似率。大多数国家的网络相似率高于边界相似率。边界角度的相似率与国家网络和运营商网络的相似率呈正相关性关系。在覆盖的全球 187 个主权国家中, 有 17 个国家相似率为 1, 这些国家的国家网络和运营商网络中网络和边界的路由器和连接完全一致。

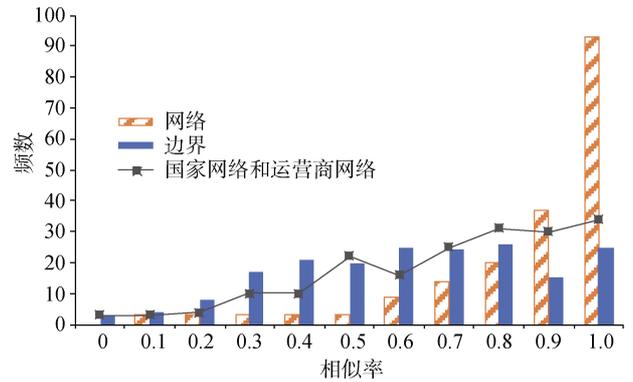


图 6 国家网络与运营商网络的相似率的频数

Figure 6 The frequency of similarity between national network and operator's network

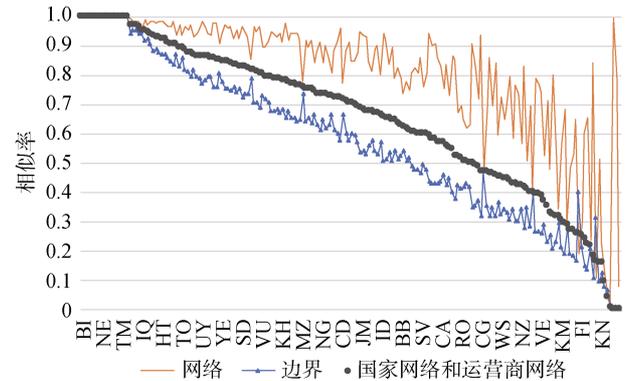


图 7 国家网络与运营商网络的相似率

Figure 7 The similarity between national network and operator's network

4.3.2 境外占比差结果

国家网络与运营商网络的境外占比差结果如图所示, 图 8 和图 9 分别表示境外占比差频数分布和各国家对应的境外占比差(图中以国家网络与运营商网络的境外占比差降序排列, 横坐标国家为每七个国家中的第一个), 国家网络和运营商网络的境外占比

差的平均值为 **0.15**, 121 个国家低于平均值, 在 0-0.1 的区间分布最多为 84 个国家; **网络境外占比差平均值为 0.14**, 129 个国家低于平均值, 在 0-0.1 的区间分布最多为 97 个国家; **边界境外占比差平均值为 0.32**, 106 个国家低于平均值, 在 0-0.1 和 0.2-0.3 区间分布多均为 30 个国家。图 9 展示了每个国家在网络和边界的境外占比差和国家网络和运营商网络的境外占比差, 大多数国家的网络境外占比差低于边界境外占比差。在国家网络与运营商网络的境外占比差降序排列中, 边界境外占比差波动较大。在覆盖的全球 187 个主权国家中, 有 17 个国家境外占比差为 0, 与相似率为 1 的国家相同。

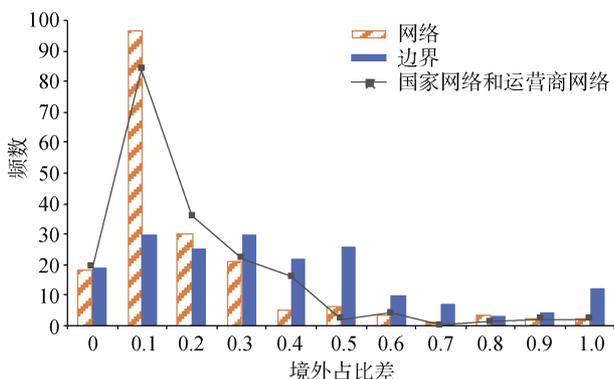


图 8 国家网络与运营商网络的境外占比差的频数
Figure 8 The frequency of external ratio difference between national network and operator's network

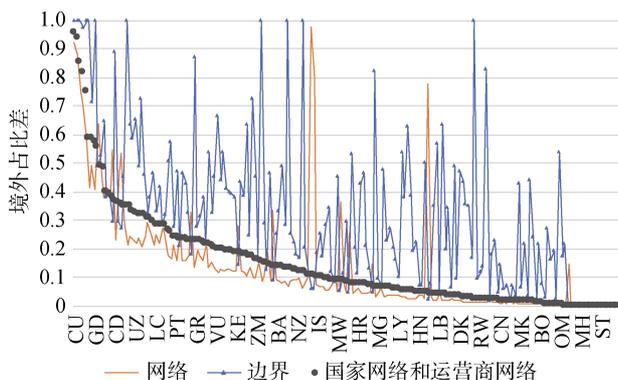


图 9 国家网络与运营商网络的境外占比差
Figure 9 The external ratio difference between national network and operator's network

4.3.3 一致性结果

每个国家的国家网络与运营商网络的相似率和境外占比差关系如图 10 所示(图中以国家网络与运营商网络的相似率降序排列, 横坐标国家为每六个国家中的第一个)。以相似率和境外占比差的平均值为基准, **50.8% 的国家是国家网络和运营商网络的相似率越高、境外占比差越低**; 5.3% 的国家是相似

率越高、境外占比差越高; 13.4% 的国家是相似率越低、境外占比差越低; 30.5% 的国家是相似率越低、境外占比差越高。整体而言, **国家网络与运营商网络的相似率越高, 境外占比差越低**。有 11.8% 的国家的境外占比差大于相似率。每个主权国家的国家网络与运营商网络的相似率和境外占比差值见附录表 1 所示。

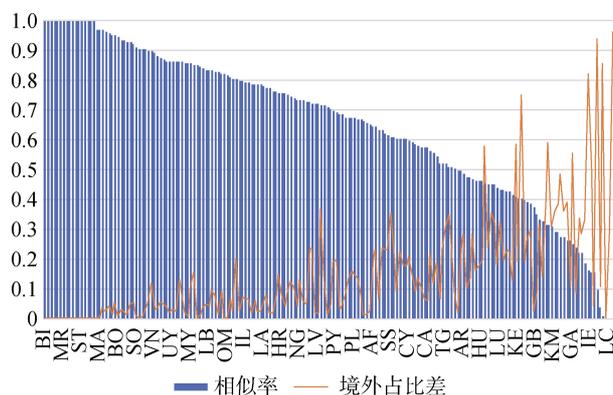


图 10 国家网络与运营商网络的一致性
Figure 10 The consistency between national network and operator's network

4.4 局限性

从相似率和境外占比差两个部分, 网络 and 边界两个角度、路由器和连接两个方面分析国家网络和运营商网络的一致性, 得到了主权国家的国家网络与运营商网络的一致性值。主权国家的国家网络和运营商网络的平均网络相似率高于 0.82 而边界相似率仅有 0.58, 大多数国家的网络相似率高于边界相似率; 平均网络境外占比差低于 0.14 而边界境外占比差高于 0.32, 大多数国家的网络境外占比差低于边界相似率。各国家的相似率越高境外占比差越低则一致性越高, 以均值为界, 50.8% 的国家一致性较高, 相似率较高且境外占比差较低。

以作者主观认知从网络和边界两个角度采用相似率和境外占比差分析国家网络和运营商网络一致性存在一定的主观性。以相同的权值综合网络和边界两个角度的指标未考虑不同国家的网络和边界的权重差异, 例如西欧国家间网络连接繁杂, 边界权重应大于网络权重。以节点度平均值归一化处理路由器和连接两个方面的指标, 未考虑不同类型路由器和连接的权重差异, 例如国家的骨干路由器和连接的权重应远大于边缘路由器和连接的权重。

5 互联网基础设施部署现状分析

以路由器的数目类比互联网基础设施的数目, 根

据国家网络和运营商网络的一致性值, 提出了三个与互联网基础设施部署现状相关的研究问题: (1)各个国家的互联网基础设施数目有何差异? (2)各个国家在互联网基础设施部署位置上有何差异? (3)各个国家的互联网基础设施所有权和管辖权有何差异?

5.1 部署规模

路由器级拓扑数据覆盖全球 187 个主权国家。各个国家的互联网基础设施的部署规模如图所示, 图 11 和图 12 分别表示国家网络和运营商网络(横坐标表示以纵坐标单位进行降序排列, 横坐标国家为每七个国家中的第一个): 以国家网络划分拓扑数据, 前 26 个国家的路由器总数占路由器总数的 90%, 仅 26 个国家的连接总数高于全球连接总数的平均值, 前 23 个国家的连接总数占连接总数的 90%。以国家运营商网络划分拓扑数据, 前 25 个国家的路由器总数占路由器总数的 90%, 仅 25 个国家运营商的路由器数目高于全球路由器数目的平均值, 21 个国家的连接总数占连接总数的 90%。

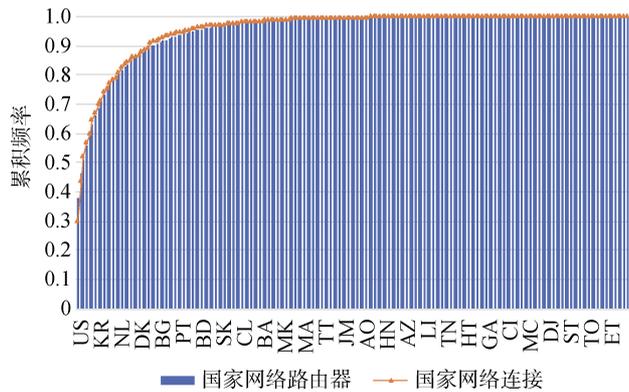


图 11 国家网络的路由器和连接数量累积分布图

Figure 11 The cumulative distribution of routers and links in national network

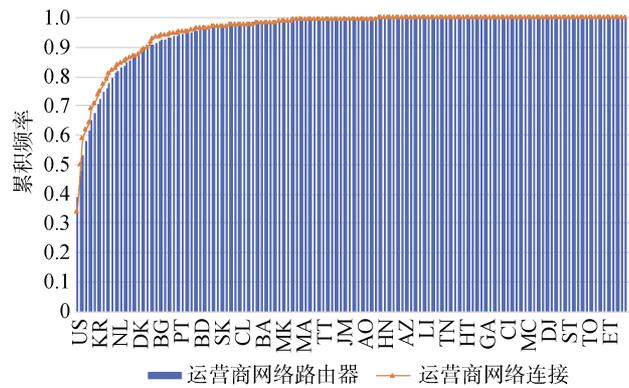


图 12 国家运营商网络的路由器和连接数量累积分布

Figure 12 The cumulative distribution of routers and links in operator's network

26 个国家拥有超过 90% 的全球互联网基础设施。少数的国家拥有着全球互联网大部分的资源。其中, 美国的运营商网络中路由器占全球路由器总数的 39.2%, 中国、德国、英国分别占 8.6%、5.2%、5.0%。

5.2 部署范围

国家运营商的互联网基础设施可能分布在多个国家, 同理, 国家领土上的互联网基础设施可能归属于多个国家。国家领土中互联网基础设施归属的国家数目与国家运营商互联网基础设施分布的国家数目如图所示, 图 13 和图 14 分别表示国家数目的频数分布和各国家对应的数目(横坐标表示国家的数目降序排列, 横坐标国家为每七个国家中的第一个)。国家网络和运营商网络中不同国家数目的频数分布较相似, 58% 的网络中包含 0-5 个不同国家。在国家网络中, 平均每个国家领土上有 9 个国家的互联网基础设施, 仅 5.9% 的国家领土上有超过 50 个国家的互联网基础设施。在国家运营商网络中, 平均每个国家运营商的互联网基础设施分布在 8.5 个国家, 仅 4.8% 的国家运营商的互联网基础设施分布在超过 30 个国家。

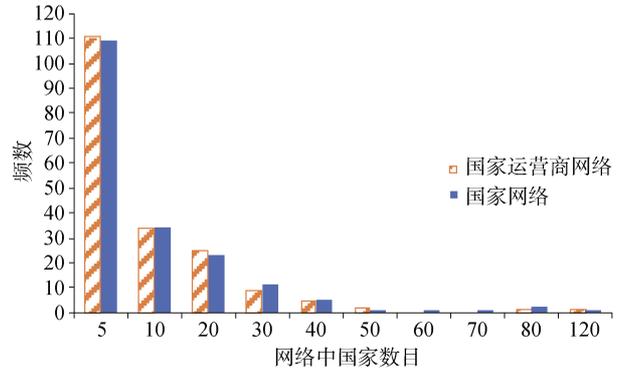


图 13 国家网络与运营商网络中不同国家数量的频数

Figure 13 The frequency of number of different nations in national network and operator's network

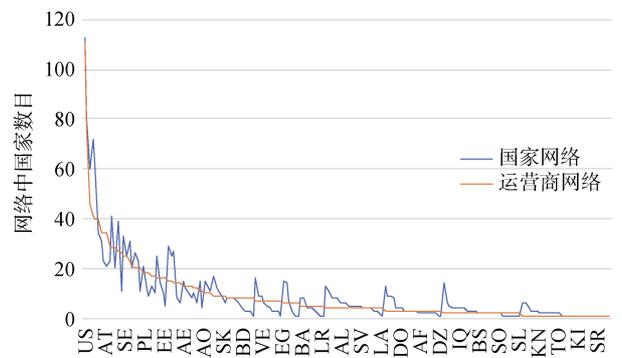


图 14 国家网络与运营商网络中不同国家数量

Figure 14 The number of different nations in national network and operator's network

极少数国家与众多国家的互联网基础设施的部署关系密切。美国领土上部署了 113 个国家的互联网基础设施,同时美国运营商的互联网基础设施分布在 111 个国家;英国领土上部署了 80 个国家的互联网基础设施,同时英国运营商的互联网基础设施分布在 76 个国家。其次为德国、荷兰、法国等与他国互联网基础设施部署关系密切。**大多数国家仅与 5 个国家以内的互联网基础设施部署关系密切。**34 个国家领土上均为本国互联网基础设施,32 个国家运营商的互联网基础设施只分布在本国,其中 17 个国家领土上仅有本国互联网基础设施且其运营商的互联网基础设施只分布在本国。

5.3 部署一致性

互联网基础设施的所有权为其所属运营商的国家,而管辖权为其部署位置的国家。基于国家网络与运营商网络的一致性将各个国家互联网基础设施的所有权和管辖权分类为一致性高、一致性高但不均衡、一致性低但均衡和一致性低。**不同国家比较来看,国家网络和运营商网络一致性越高,国家运营商网络的地理边界越明显。**

以各个国家相似率和境外占比差的平均值为界:

(1) 当国家网络和运营商网络相似率高且境外占比差低时,国家互联网基础设施所有权和管辖权的一致性高,国家对。有 95 个国家的互联网基础设施的相似率高于 66.0%,境外占比差低于 14.8%,其中 17 个国家的互联网基础设施的所有权和管辖权完全一致。在网络和边界角度分析一致性时,埃及、孟加拉国、越南、泰国等国家的网络中互联网基础设施的一致率和网络边界的一致率均很高;而俄罗斯、丹麦的网络边界中互联网基础设施的一致率较低。在境外占比差分析部署均衡性时,中国和以色列的互联网基础设施分布在他国的数目远多于他国部署在本国领土的数目;巴西、韩国、印度、土耳其相反。

中国、俄罗斯、韩国和印度的国家互联网基础设施的所有权和管辖权一致性高,国家网络和运营商网络中路由器和链接相同的数目多,边界路由器和链接相同的数目同样较多。中国运营商网络分布在他国的路由器数目远大于其他国家部署在本土的数目,韩国和印度相反。俄罗斯运营商网络分布在其他国家的路由器略小于其他国家部署在本土的数目。

(2) 当国家网络和运营商网络相似率高但境外占比差高时,国家互联网基础设施所有权和管辖权的一致性高但不均衡。有 10 个国家的互联网基础设施的一致率高但不均衡,其中希腊和波兰等国家分

布在他国的互联网基础设施的数目远少于他国部署在本国领土的数目。

波兰的国家互联网基础设施的所有权和管辖权一致性较高,其他国家部署在波兰领土上的路由器数目远大于波兰运营商分布在其他国家的数目。

(3) 当国家网络和运营商网络的相似率低但境外占比差低时,国家互联网基础设施所有权和管辖权的一致性低但均衡。有 25 个国家的互联网基础设施的一致性低但均衡,特别是国家的边界互联网基础设施一致率低,如美国网络中互联网基础设施的相似率为 91.1%但边界的相似率仅为 34.3%;日本、英国、德国、墨西哥、南非、乌克兰、挪威、新西兰、印度及西亚、芬兰等国家的网络和边界相似率的差值达 30%以上。

美国、英国和日本的国家互联网基础设施的所有权和管辖权一致性较低,美国和英国运营商网络分布在他国的路由器数目远大于其他国家部署在本土的数目,日本相反。美国网络边界的路由器与运营商网络边界的路由器数目相似,但相同的数量较少;英国和日本网络边界的路由器数目远大于运营商网络边界的路由器数目,相同的数目较少。

(4) 当国家网络和运营商网络相似率低且境外占比差高时,国家互联网基础设施所有权和管辖权的一致性差。有 57 个国家的互联网基础设施的一致率低,如瑞士、新加坡、法国、瑞典、西班牙、澳大利亚网络中互联网基础设施的相似率与边界的相似率差值大于 44.2%,并且互联网基础设施分布在他国的数目远小于他国在本国部署的数目;其中,22 个国家的互联网基础设施的相似率甚至低于境外占比差,如爱尔兰、荷兰、奥地利等国家的互联网基础设施分布在他国的数目与他国部署在本国的数目相对较大且数目间差值较大。

法国和奥地利的国家互联网基础设施的所有权和管辖权一致性低,法国运营商网络分布在他国的路由器数目远大于其他国家部署在本土的数目,奥地利相反。奥地利网络边界的路由器与运营商网络边界的路由器数目相似,但相同的数量及少;法国网络边界的路由器数目远大于运营商网络边界的路由器数目,相同的数目较少。

5.4 部署影响

对互 3 网而言,美国建设因特网在先,各国接入美国因特网在后,从而形成了由美国主导的集中管理模式^[24]。基于 CAIDA 数据标注基础设施的 187 个主权国家中,少数的国家拥有着全球互联网大部分的资源,与众多国家的互联网基础设施关

联密切。美国拥有 39.2% 的互联网资源部署在 111 个国家, 而 109 个国家的互联网基础设施仅部署在 5 个以内的国家。

随着网络技术快速发展, 各国运营商在海外设立分支机构部署互联网基础设施。管辖权和所有权不同的互联网基础设施在数据跨境流通和安全保护等方面对各国经济、外交和法律产生了深远影响。数据跨境流通和便捷的国际数字贸易发展直接影响了经济发展。数据自由流动和数据安全保护不仅涉及个人的数据与隐私安全, 更关乎国家安全与政治安全。数据治理成为全球各国争取数字经济规则制定权的关键议题。

关键基础设施作为国家重要战略资源被列为主要攻击目标。自 2022 年俄乌冲突爆发以来, 均发动了针对对方基础设施和信息系统的网络战, 一些大型国际互联网企业也纷纷介入, 破坏了国家间的外交信任。有国家背景的黑客组织、APT 组织等对关键信息基础设施发动网络攻击, 或通过植入后门操纵国际信息安全标准进行数字监听等行径, 不仅动摇了互联网的技术信任基础, 对全球国际关系生态环境造成极为恶劣影响。

网络空间军事化态势进一步凸显, 关键基础设施成为网络战的主战场。于我国而言, 根据国家互联网应急中心的数据, 境外国际级攻击组织对我国进行长达 11 年的网络攻击, 甚至将网络意识形态渗透作为国家“软实力”扩张的重要策略, 使我国网络空间主流意识形态话语权面临严峻挑战^[25]。

随着数字化进程在各行业领域的逐步推进, 传统国家安全产业领域面临的网络安全风险加剧, 我国应以关键信息基础设施为重点, 加快网络安全防御体系建设, 构建配套的技术平台, 出台系统化的法律法规。如近期实施的《信息安全技术 关键信息基础设施安全保护要求》有助于进一步提升国家关键信息基础设施安全保障能力和水平, 推动我国网络强国战略实施、数字经济的健康良性发展^[26]; 为满足日益增长的个人信息出境需要, 保护个人信息权益出台的《个人信息出境标准合同办法》是网络空间综合治理机制的重要创新^[27]。

在和平环境下, 居安思危。对外: 在参与网络空间国际治理体系改革、构建国内网络治理体系的过程中, 必须坚持网络主权, 明确国内互联网发展路径、格局, 保证国家网络基础设施建设、运营、维护和使用网络。对内: 加强内部自身能力建设, 构建完整的网络作战架构, 寻求组织网络攻击和在战时保护关键网络的工业技术, 以保护天基网络、地面站和基础设施。

5.5 局限性

当排除政治因素, 仅以数据驱动分析国家网络和运营商网络的一致性, 并以相似率和境外占比差值大小比较国家的互联网基础设施部署间的差异本身具有局限性。以路由器数目量化互联网基础设施的部署规模, 以路由器的地理国家和运营商归属国家判定互联网基础设施管辖权和所有权, 未区分不同互联网基础设施的权重, 如骨干网路由器和普通路由器间的区别。

原始数据中部分国家的国家网络和运营商网络中路由器仅有几十个, 如尼泊尔、也门、古巴、埃塞俄比亚、几内亚等, 仅以路由器级拓扑数据分析互联网基础设施的现状误差较大。

6 结论

本文构建了分析国家网络和运营商网络一致性的模型。首先, 以 CAIDA 公开路由器级网络拓扑数据为基础, 采用被动 IP 地理定位、运营商映射和基于路由器约束的启发式方法标注路由器的地理位置和运营商国家, 获得了一份覆盖率为 99.12%, 经 IXP 验证标注地理国家的准确率为 96.47%, 运营商国家的准确率为 75.08% 的路由器的国家级“管辖权-所有权”的数据集。然后, 从网络和边界两个角度分析国家网络和运营商网络的部署规模与不同国家互联互通; 从相似率和境外占比差两部分提出国家网络和运营商网络的境内一致性和境外差异性的指标; 从路由器和连接两个方面量化国家网络与运营商网络的一致性值。获得了 187 个主权国家的国家网络和运营商网络相似率和境外占比差值。最后, 以路由器数目比作互联网基础设施数目, 研究全球互联网基础设施的部署现状。本文发现, 仅有 26 个国家的互联网基础设施总数目已经超过 90% 的全球各国家互联网基础设施数目之和, 互联网基础设施部署在几十个国家, 如美国分布在 111 个国家, 而大多数国家拥有极少的互联网资源, 其中 109 个国家的互联网基础设施仅部署在 5 个国家以内。各个国家的国家网络和运营商网络中互联网基础设施的一致性比边界互联网基础设施一致性高, 部分国家分布在他国的互联网基础设施比他国部署在本国的多, 如中国、以色列、美国、英国和挪威等而日本、墨西哥、芬兰、德国等国家恰好相反。

在未来的研究中, 将从更多角度分析国家网络和运营商网络的一致性, 能够针对不同国家的网络特征设定不同指标和权重, 能够识别分析各国互联网基础设施。重点考虑当特定国家网络中断时, 对其他国家网络的影响。

参考文献

- [1] Russia just brought in a law to try to disconnect its internet from the rest of the world. <https://www.cnbc.com/2019/11/01/russia-controversial-sovereign-internet-law-goes-into-force.html>. Nov. 2019.
- [2] The clean network. <https://2017-2021.state.gov/the-clean-network/index.html>. Aug. 2020.
- [3] KA-SAT Network cyber attack overview. <https://news.viasat.com/blog/corporate/ka-sat-network-cyber-attack-overview>. Feb. 2022.
- [4] A new iron curtain is descending across Russia's Internet. <https://www.washingtonpost.com/technology/2022/03/04/russia-ukraine-internet-cogent-cutoff/>. Mar. 2022.
- [5] Mintsifra: Disabling MegaFon and Rostelecom from LINX will not affect Internet speed. <https://time.news/mintsifra-disabling-megafon-and-rostelecom-from-linx-will-not-affect-internet-speed/>. Mar. 2022.
- [6] Internet shutdowns. <https://pulse.internetsociety.org/shutdowns>. Aug. 2022.
- [7] Motamedi R, Rejaie R, Willinger W. A Survey of Techniques for Internet Topology Discovery[J]. *IEEE Communications Surveys & Tutorials*, 2015, 17(2): 1044-1065.
- [8] Zhang H L, Fang B X, Hu M Z, et al. A Survey on Internet Measurement and Analysis[J]. *Journal of Software*, 2003, 14(1): 110-116.
(张宏莉, 方滨兴, 胡铭曾, 等. Internet 测量与分析综述[J]. *软件学报*, 2003, 14(1): 110-116.)
- [9] Zhang Y, Zhang H L, Fang B X. A Survey on Internet Topology Modeling[J]. *Journal of Software*, 2004, 15(8): 1220-1226.
(张宇, 张宏莉, 方滨兴. Internet 拓扑建模综述[J]. *软件学报*, 2004, 15(8): 1220-1226.)
- [10] Zhang Y, Fang B X, Zhang H L. Chinese IP-Level Network Topology Measurement and Analysis[J]. *Journal on Communications*, 2007, 28(12): 96-101.
(张宇, 方滨兴, 张宏莉. 中国 IP 级网络拓扑测量与分析[J]. *通信学报*, 2007, 28(12): 96-101.)
- [11] Luckie M, Huffaker B, Marder A, et al. Learning to Extract Geographic Information from Internet Router Hostnames[C]. *The 17th International Conference on Emerging Networking Experiments and Technologies*, 2021: 440-453.
- [12] Luckie M, Marder A, Fletcher M, et al. Learning to Extract and Use ASNs in Hostnames[C]. *The ACM Internet Measurement Conference*, 2020: 386-392.
- [13] Scheitle Q, Gasser O, Sattler P, et al. HLOC: Hints-Based Geolocation Leveraging Multiple Measurement Frameworks[C]. *2017 Network Traffic Measurement and Analysis Conference*, 2017: 1-9.
- [14] Zhu J Y, Zhang Y, Zeng L W, et al. Geolocation for Multi-Interface Routers[J]. *Journal of Cyber Security*, 2018, 3(4): 15-24.
(朱金玉, 张宇, 曾良伟, 等. 一种多接口路由器地理定位方法[J]. *信息安全学报*, 2018, 3(4): 15-24.)
- [15] Zhu J Y, Zhang Y, Zeng L W, et al. Method of Discovering Regional Network Border[J]. *Journal of Software*, 2023, 34(3): 1512-1522.
(朱金玉, 张宇, 曾良伟, 等. 地区网络边界发现方法[J]. *软件学报*, 2023, 34(3): 1512-1522.)
- [16] Giotsas V, Smaragdakis G, Huffaker B, et al. Mapping Peering Interconnections to a Facility[C]. *The 11th ACM Conference on Emerging Networking Experiments and Technologies*, 2015: 1-13.
- [17] Ziv M, Izhikevich L, Ruth K, et al. ASdb[C]. *The 21st ACM Internet Measurement Conference*, 2021: 703-719.
- [18] Carisimo E, Gamero-Garrido A, Snoeren A C, et al. Identifying ASes of State-Owned Internet Operators[C]. *The 21st ACM Internet Measurement Conference*, 2021: 687-702.
- [19] CAIDA ITDK. <https://www.caida.org/catalog/datasets/internet-topology-data-kit/>. Oct. 2022.
- [20] MAXMIND. <https://www.maxmind.com/en/geoip2-databases>. Oct. 2022.
- [21] IP2AS. <https://bgpview.io/>. Aug. 2020.
- [22] AS2Country. <https://www.cidr-report.org/as2.0/autnums.html>. Aug. 2020.
- [23] PeeringDB. <https://www.peeringdb.com/>. Oct. 2022.
- [24] Fang B X. *On the sovereignty of cyberspace*[M]. Beijing: Science Press, 2017: 464.
(方滨兴. 论网络空间主权[M]. 北京: 科学出版社, 2017: 464.)
- [25] Lai X X, Zhao Z Y, Yuan W P. Achievements, Challenges and Optimization Path of China's Network Governance System[J]. *China Information Security*, 2023(4): 64-67.
(赖星星, 赵志云, 袁卫平. 我国网络治理体系的成就、挑战与优化路径[J]. *中国信息安全*, 2023(4): 64-67.)
- [26] 《信息安全技术 关键信息基础设施安全保护要求》. https://www.gov.cn/xinwen/2022-11/07/content_5725199.htm. Nov. 2022.
- [27] 《个人信息出境标准合同办法》公布. https://www.sac.gov.cn/xw/mtjj/art/2023/art_9e3515bdc1ab4aea901d7cab9eb35b8d.html. Feb. 2023.



朱金玉 于 2015 年在四川大学计算机科学与技术专业获得学士学位。现在哈尔滨工业大学网络空间安全学院攻读博士学位。CCF 会员。研究领域为网络空间测绘, 互联网关键资源安全, IP 地理定位。Email: zhujinyu@hit.edu.cn



王宇楠 于 2022 年在哈尔滨工业大学网络空间安全学院获得硕士学位。研究领域为互联网关键资源安全, 网络拓扑测量体系结构。Email: 1273652804@qq.com



张宇 男, 博士, 哈尔滨工业大学网络空间安全学院副教授、博士生导师, 研究领域为互联网关键资源安全, 网络拓扑测量, 未来网络体系结构。Email: yuzhang@hit.edu.cn



张宏莉 女, 博士, 哈尔滨工业大学网络空间安全学院教授、博士生导师, 研究领域为网络安全, 网络测量和网络计算。Email: zhanghongli@hit.edu.cn



方滨兴 男, 中国工程院院士, 哈尔滨工业大学网络空间安全学院教授、博士生导师, 研究领域为大数据、计算机网络与信息安全。Email: fangbx@cae.cn

附 录

附录表 1 拓扑数据覆盖的主权国家的国家网络和运营商网络的相似率和境外占比差值

Appendix Table 1 Similarity and external ratio difference values of national network and operator's network of sovereign nations covered by topological data

Nation	JS	JD												
BI	1.00	0.00	SZ	0.87	0.05	TR	0.76	0.08	SK	0.61	0.20	SE	0.42	0.22
CV	1.00	0.00	LY	0.87	0.06	HR	0.76	0.08	SS	0.61	0.35	PA	0.42	0.23
FM	1.00	0.00	BT	0.86	0.01	MZ	0.76	0.15	NO	0.60	0.09	NZ	0.41	0.12
GW	1.00	0.00	CN	0.86	0.02	LT	0.75	0.04	TZ	0.60	0.18	TJ	0.41	0.59
KI	1.00	0.00	GM	0.86	0.02	BG	0.75	0.09	SV	0.60	0.18	KE	0.40	0.19
MH	1.00	0.00	UY	0.86	0.03	BF	0.74	0.10	CY	0.60	0.21	CF	0.40	0.75
MR	1.00	0.00	SI	0.86	0.03	AL	0.74	0.12	IT	0.60	0.22	FR	0.39	0.19
NE	1.00	0.00	IN	0.86	0.04	SB	0.73	0.04	GY	0.59	0.15	SG	0.39	0.26
NR	1.00	0.00	MD	0.86	0.09	ME	0.73	0.05	MN	0.58	0.09	CH	0.39	0.29
PW	1.00	0.00	MY	0.86	0.10	DO	0.73	0.06	JP	0.58	0.13	VE	0.37	0.02
SM	1.00	0.00	GQ	0.86	0.13	NG	0.73	0.13	CA	0.57	0.06	GB	0.35	0.09
SR	1.00	0.00	GE	0.85	0.00	GR	0.73	0.23	MG	0.57	0.07	GT	0.33	0.32
ST	1.00	0.00	MK	0.85	0.02	SA	0.72	0.02	ZA	0.57	0.11	EE	0.32	0.14
SY	1.00	0.00	YE	0.85	0.16	LV	0.72	0.02	BW	0.56	0.22	NL	0.31	0.31
TM	1.00	0.00	PK	0.84	0.04	GN	0.72	0.23	NI	0.55	0.12	CR	0.31	0.40
TT	1.00	0.00	RS	0.83	0.01	CD	0.72	0.37	EC	0.55	0.19	BZ	0.31	0.59
ET	1.00	0.00	LB	0.83	0.04	RU	0.71	0.01	MX	0.52	0.07	KM	0.29	0.36
EG	0.97	0.00	CI	0.83	0.05	KZ	0.71	0.13	SC	0.52	0.23	ZW	0.29	0.38
MA	0.97	0.01	SD	0.83	0.08	HN	0.70	0.05	TG	0.52	0.31	PE	0.27	0.36
NP	0.97	0.03	TN	0.83	0.09	GH	0.70	0.19	DJ	0.51	0.27	AO	0.27	0.48
SL	0.96	0.03	TD	0.82	0.00	DK	0.69	0.03	RO	0.51	0.35	KW	0.26	0.11
CM	0.95	0.01	OM	0.82	0.00	PY	0.69	0.20	US	0.50	0.02	DZ	0.26	0.39
IQ	0.95	0.04	PG	0.82	0.10	JM	0.68	0.04	MW	0.50	0.09	GA	0.25	0.56
AZ	0.95	0.05	BS	0.81	0.06	FJ	0.68	0.08	PT	0.50	0.24	FI	0.24	0.08
BO	0.94	0.01	RW	0.80	0.02	BA	0.68	0.14	AR	0.49	0.28	MU	0.22	0.28
BD	0.93	0.01	BY	0.80	0.03	NA	0.67	0.08	IS	0.47	0.10	AT	0.22	0.34
QA	0.93	0.02	JO	0.80	0.07	AE	0.67	0.14	CG	0.47	0.14	UZ	0.18	0.33
TL	0.93	0.03	VU	0.80	0.20	PL	0.67	0.15	CO	0.47	0.29	LR	0.16	0.05
HT	0.93	0.05	BH	0.79	0.02	ZM	0.67	0.16	ML	0.46	0.16	IE	0.16	0.49
KG	0.92	0.05	IL	0.79	0.06	ID	0.66	0.02	CZ	0.46	0.18	SN	0.16	0.82
IR	0.91	0.00	AM	0.79	0.06	BN	0.65	0.01	HU	0.46	0.20	KN	0.09	0.94
SO	0.91	0.01	MT	0.79	0.07	DE	0.65	0.03	CL	0.45	0.24	DM	0.04	0.11
TH	0.91	0.01	KR	0.78	0.02	AF	0.65	0.19	BE	0.45	0.32	AG	0.01	0.86
LS	0.90	0.06	KH	0.78	0.02	UG	0.64	0.23	WS	0.45	0.35	LC	0.00	0.29
LK	0.90	0.03	LA	0.78	0.06	UA	0.63	0.07	GD	0.45	0.58	CU	0.00	0.96
VN	0.89	0.04	MV	0.77	0.02	LI	0.63	0.24	AU	0.44	0.18	AVE	0.66	0.15
TO	0.89	0.12	BR	0.77	0.08	BB	0.62	0.23	ES	0.43	0.20			
MM	0.88	0.03	PH	0.76	0.02	MC	0.62	0.23	LU	0.43	0.32			