

基于拓扑分析的区域级网络抗毁性研究

安常青^{1,2}, 刘玉家¹, 王会¹, 郑志延¹, 喻涛¹, 王继龙^{1,2,3}

(1. 清华大学网络科学与网络空间研究院北京信息科学与技术国家研究中心, 北京 100084;
2. 鹏城实验室, 广东 深圳 518000; 3. 清华大学奇安信联合研究中心, 北京 100084)

摘要: 为了衡量区域受到不同攻击的整体影响, 捕捉区域网络拓扑之间的抗毁性差异, 提出一种区域抗毁性评估算法。从区域内部和区域外部出发, 分析区域内部自身抗毁性和区域之间通信拓扑的抗毁性。多方获取测量数据并校验以构建拓扑关系图, 实现区域间链接的预测作为拓扑数据的补充。基于分层概率采样, 多次模拟破坏以逼近真实情况下区域的受破坏情况。实现显著性检验器, 分别从整体水平和特殊薄弱点衡量区域受破坏的影响, 发现区域之间受破坏情况的差异, 计算区域抗毁性排名。最终给出 48 个区域的排名和聚类结果。

关键词: 区域; 拓扑; 抗毁性; 安全性; 显著性差异

中图分类号: TP393

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021179

Research on the invulnerability of regional network based on topology analysis

AN Changqing^{1,2}, LIU Yujia¹, WANG Hui¹, ZHENG Zhiyan¹, YU Tao¹, WANG Jilong^{1,2,3}

1. Institute for Network Sciences and Cyberspace, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

2. Peng Cheng Laboratory, Shenzhen 518000, China

3. Tsinghua University and QI-ANXIN Joint Research Center, Beijing 100084, China

Abstract: To measure the overall impact of different attacks a network region may encounter, an evaluation algorithm can be used to capture the difference in invulnerability of regional network topologies. Specifically, the vulnerability of regional network and that of connection topology between regions both internally and externally was analyzed. To find predictions between regions as supplemental topology data and create the Internet topology, data would be measured in various ways and verified. Using stratified probability sampling, multiple damage simulations could predict the worst-case scenarios in real conditions. Damage was simulated multiple times based on stratified probability sampling to approximate the damage situation of the area under real conditions. A significance tester was realized to measure the impact of damage from the overall level and special weak points respectively, discover the difference of the damage between the areas, and calculate the regional invulnerability ranking. Finally, the ranking and clustering results of forty-eight regions were given.

Keywords: region, topology, invulnerability, security, significant difference

1 引言

全球互联网的重要性日益提高, 已成为人们生活各个方面所依赖的关键基础设施之一, 网络的安全性是保证正常通信的基础。

对于网络拓扑, 目前区域或地区受到多方面的安全威胁。首先, 各种类型的恶意行为者针对网络拓扑弱点破坏、拦截互联网流量^[1-2], 通过对网络进行攻击, 实现各种政治、经济等目标。

其次, 已有研究发现部分区域或地区具有与互

收稿日期: 2021-06-09; 修回日期: 2021-08-23

基金项目: 国家自然科学基金资助项目 (No.62072269); 国家重点研发计划基金资助项目 (No.2020YFE0200500)

Foundation Items: The National Natural Science Foundation of China (No.62072269), The National Key Research and Development Program of China (No.2020YFE0200500)

联网断开的可能, Dyn 针对与国外互联网服务提供商 (ISP, Internet service provider) 有直接联系的国内互联网服务提供商进行分析, 发现只有 30 个区域或地区没有与全球互联网脱钩的风险。目前有专业的机构进行网络断网事件的监测并实时报告, 如千眼、互联网健康报告。

除上述特定的恶意攻击外, 自然灾害对网络设施的威胁也是巨大的^[3], 通常会影响到特定地理区域的所有目标, 即使是很发达的区域, 也会由于自然灾害的原因受到影响, 例如 2012 年美国东海岸桑迪飓风导致的断网^[4]。

学术界在网络拓扑弹性、安全性及抗毁性方面持续有相关研究。早期工作大致可总结为两类, 针对网络拓扑建模并对网络特点开展相关分析或是基于早期测量数据进行抗毁性分析。Albert 等^[5]通过随机性删除网络节点, 分析了复杂网络的错误和攻击容忍度, 发现互联网具有“意外程度的稳健性”, 甚至不切实际的高故障率也不会影响节点的通信能力。一些有关互联网拓扑弹性的工作^[6]基于简化的拓扑图, 没有加入路由策略的限制, 从图论的角度对拓扑进行分析。

之后研究扩展到具有路由策略的实际拓扑, 进一步分析链路故障的位置如何影响互联网。Wu 等^[7]和 Dolev 等^[8]从破坏性角度分析网络拓扑的弹性。Dombrowski 等^[9]从图论的角度分析云服务的拓扑连通性, 模拟了随机和有针对性的节点攻击, 并评估了云服务的相应漏洞。Omer 等^[10]提出了用于测量网络弹性的模型, 定义了网络中断前后的传输价值来衡量网络的基本弹性, 基于现有的容量和流量信息, 分析光缆网络的全局弹性。

学术界逐步由关注网络整体变为关注于网络中区域本身或某些重要组件的安全性。近期研究工作更多地集中在对区域整体以及区域间直连自治系统 (AS, autonomous system) 链路的分析, 即边界 AS 链接。Leyba 等^[11]定义国家关键点潜力 (NCP, national chokepoint potential), 通过 NCP 评估边界 AS 在区域拓扑的影响力。Alexander 研究传输运营商在区域层面的影响力, 主要是每个区域中最具影响力的 AS, 以及那些严重依赖传输 AS 的区域。CAIDA 项目 MapKit 研究互联网拓扑对区域/地区级连接中断的敏感性。项目提案中提出构建网络的多层拓扑图, 涵盖网络层、链路层、物理层, 识别并量化区域对攻击的敏感性。针对区域的拓扑结构,

上述研究并未定量分析区域的抗毁性差异。

本文主要贡献如下, 提出了区域抗毁性评估方法, 定义了抗毁性度量, 分别从内部拓扑网络和外部通信拓扑 2 个方面对区域的抗毁性度量差异进行评估。基于显著性检验, 分别从网络整体水平和网络特殊薄弱点 2 个方面对抗毁性度量排名, 该方法具有普适性, 能给发现区域之间拓扑的本质差异。数据层面考虑到探测点的局限性, 进行区域之间直连链路预测, 完善拓扑数据。

2 定义

本节描述本文涉及的主要定义及相应的含义。

1) AS 抗毁性: 一个事件发生后 AS 与其他节点间通信情况的变化称为 AS 对一个事件的抗毁性。AS 针对多次破坏的抗毁性形成一维向量称为 AS 的抗毁性样本。

2) 区域抗毁性: 区域抗毁性是对区域内 AS 抗毁性的加权综合度量。

3) 路由影响力: 描述其他节点需要经过该 AS 进行转发的依赖程度, 即 AS 在区域的路由影响。

4) 资源权重: 描述节点在区域中持有的资源比例, 即占该区域所有资源的比例。节点资源权重越高, 被破坏后对区域的影响越大, 用于量化破坏事件的影响。本文从用户数量、域名等多个角度分析节点的资源权重, 形成不同视角下的抗毁力评估。

6) 边界 AS: 与其他区域有直接的拓扑链接。

7) 内部 AS: 没有与其他区域的直接拓扑链接。

8) 边界 AS 链接: 不同区域边界 AS 之间的链接。

下面, 详细介绍节点抗毁性、路由影响力和区域抗毁性度量的含义和计算方法。

2.1 节点抗毁性

事件发生后 AS 与其他节点间连通情况的变化称为节点对一次事件的抗毁性。针对多次破坏的抗毁性形成一维向量, 即节点的抗毁性样本, 对样本的综合分析形成节点抗毁性度量。本节重点介绍节点抗毁性样本的定义。

2.1.1 单次事件抗毁性

互联网的核心是“互连”, 学术界定义网络破坏的度量中, 常用且可解释性好的就是连通性, 可以直观反映破坏后节点的通信情况。本文使用连通性来表征破坏的影响, 即发生破坏性事件后, 网络间节点对的连通性变化程度。

对于区域中的某个 AS, 一次破坏事件可以使

其与某些节点无法连通，而区域内各个节点所持有资源的比例不同，如对于 AS 节点，从用户角度，拥有更多用户的 AS 资源比例更高。

破坏事件影响定义为破坏后该区域可与其连通的节点资源之和与该区域所有节点资源之和的比值，即

$$e_{ix} = \frac{\sum_{b \in B} r_b}{\sum_{o \in O} r_o} \quad (1)$$

其中， e_{ix} 表示对于节点 i ，第 x 次破坏事件的破坏影响； O 表示区域节点集合； B 表示这次破坏事件后可以与节点 i 通信的节点集合； r 表示某个节点的资源，如用户数量。

本节从不同角度出發，实现了多种节点资源 r 的定义。在区域内部，关注各个 AS 节点的连通，在 4.2.1 节定义 AS 节点的资源。在区域外部，关注各个区域之间是否连通，在 4.3.1 节定义区域节点的资源。

2.1.2 抗毁性采样

对于每个节点 i ，模拟破坏事件集合为 $[f_{i1}, f_{i2}, \dots, f_{im}]$ ，计算节点 i 面对一次破坏事件 f_{iq} 对应的单次事件抗毁性为 e_{iq} ，可以得到节点 i 的抗毁性采样 $E_i = [e_{i1}, e_{i2}, \dots, e_{im}]$ 。

下面，只讨论破坏节点的情况，破坏链接的情况类似。区域内节点数量为 n ，假设破坏一个节点的发生概率为 p ，则破坏 x 个节点的概率为 p^x 。总事件 $A = \{A_1, A_2, \dots, A_n\}$ ，其中事件 A_x 为 x 个节点被同时破坏，事件 A_x 发生的期望为 $y = C(n, x)p^x$ ，那么对于总事件 A ，所有破坏事件发生的期望为 $\sum_{x=1}^n C(n, x)p^x$ 。

图 1 为 $y = C(n, x)p^x$ 的分布。从图 1 中可以看到，首先随着破坏数量 x 的增大，实际破坏事件的发生期望 y 也急剧增多， $x=5$ 前就达到最高点，之后 y 迅速减小。当 $x \geq 10$ 时， y 均小于 0.1；当 $x \leq 6$ 时，期望值 $\sum_{x=1}^6 C(n, x)p^x$ 是所有破坏数量下总期望值

$\sum_{x=1}^n C(n, x)p^x$ 的 90%，而 $x \leq 8$ 时达到 99.9%。

本文用分层概率采样的方法，破坏节点数量 x 不同时采样概率不同，当破坏数量较大时，实际发生次数 y 很小，采样概率设为 0。

2.2 路由影响力

路由影响力衡量具体的 AS 节点在某个区域中为其他 AS 提供的路由服务，即其他的路由路径有多大比例经过该 AS。

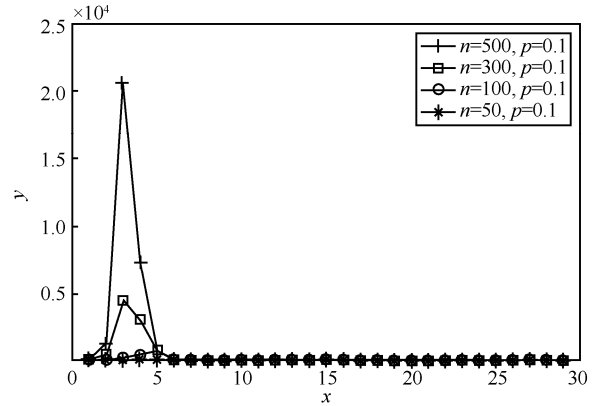


图 1 $y = C(n, x)p^x$ 分布

理想情况下，位于区域 c 的 AS_b 对于内部某个 AS_o 的路由影响力可定义为 $P(AS_b, AS_o)/r(AS_o)$ ，其中 $P(AS_b, AS_o)$ 计算 AS_o 路由经过 AS_b 的次数， $r(AS_o)$ 为 AS_o 实际路由的次数。

但是，BGP 数据收集严重偏向探测点看到的数据。由于探测点在区域之间分布不均匀，而且许多区域和大多数 AS 都没有设置探测点，获得的数据均为通往探测点的路径进行过度采样的结果。为了度量 AS 在区域内部的实际影响力，需要计算其在实际路径中的出现比例，而探测点自身的有限性和偏见性对该结果有较大影响^[12]。

本文设计了过滤器，通过对探测点进行过滤，减轻计算结果受到探测点自身局限的影响。

位于区域 c 的 AS_b 路由影响力 $w_c(AS_b)$ 定义为

$$w_c(AS_b) = \sum_{o \in O} \theta_o UR_c(AS_b, AS_o) \quad (2)$$

$$UR_c(AS_b, AS_o) = f_{vp}(\mathbf{R}) \quad (3)$$

$$R[i] = \frac{P_i(AS_b, AS_o)}{r_i(AS_o)} \quad (4)$$

$$f_{vp}(\mathbf{R}) = \frac{1}{n - 2\alpha n} \sum_{i=an+1}^{n-\alpha n} R_i \quad (5)$$

其中， O 表示区域 c 内所有 AS 集合， $UR_c(AS_b, AS_o)$ 表示 AS_b 对 AS_o 的用户影响力， θ_o 表示区域 c 中 AS_o 的用户比例。以用户比例作为权重，总路由影响力为 AS_b 对区域内所有 AS 用户影响力的加权和。式(3)计算用户影响力 UR，其中 \mathbf{R} 为一维向量，存储所有探测点实际观测到 AS_o 经过 AS_b 的比例。式(4)计算观测到经过 AS_o 的所有路径中， AS_b 经过 AS_o 的路径比例，其中 $P_i(AS_b, AS_o)$ 为探测点 i 观测到的 AS_o 经过 AS_b 的路径数量， $r_i(AS_o)$ 为探测点 i 观测到 AS_o 的数量。式(5)中， \mathbf{R} 存储所有探测点的数据，

并按照 $r_i(AS_o)$ 排序，函数 $f_{vp}(R)$ 为探测点过滤器，过滤距离 AS_b 过近和过远的探测点。

2.3 区域抗毁性

本文通过多次模拟破坏事件，获得区域抗毁性采样。设计显著性检验器，比较不同区域抗毁性采样的差异，得到各个区域的抗毁性度量。

图 2 描述了区域抗毁性的计算流程。对于每个区域，选择重要的 AS 集合，并计算每个 AS 的抗毁性采样，在 4.2.2 节和 4.3.2 节分别说明 AS 的选择标准。通过加权上采样，将区域内 AS 的抗毁性采样转换为区域的抗毁性加权采样。最后通过显著性检验器，比较不同区域抗毁性加权采样的差异，得到区域抗毁性度量。

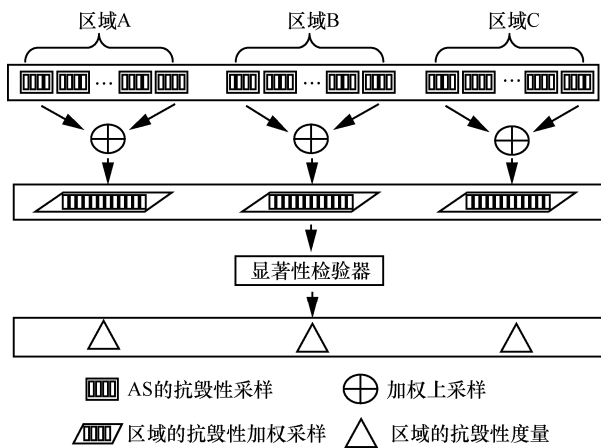


图 2 区域抗毁性计算流程

下面分别讲述加权上采样和显著性检验器的实现，然后通过采样实验和方法分析说明抗毁性计算方法的合理性。

2.3.1 加权上采样

通过加权上采样，获得区域的抗毁性采样，输入为区域中 AS 的抗毁性采样，输出为区域的抗毁性加权采样。

将 AS 的路由影响力作为该 AS 对应的采样权重。AS 的路由影响力越高，区域内依赖其通信的 AS 数量越多，且实际路由中需要经 AS 进行路由转发的次数越高，AS 的抗毁性直接影响这些依赖它的 AS 的连通性。

加权上采样思路如下，对于每个区域，输入为 $[E_1, E_2, \dots, E_m]$ 和 $[w_1, w_2, \dots, w_m]$ ，其中 E_i 和 w_i 分别表示第 i 个 AS 的抗毁性采样和采样权重，获得输入采样权重的最小值 w_{\min} ，抗毁性采样结果 E_i 的采样次数为 $\lfloor w_i / w_{\min} \rfloor$ ，最后该区域抗毁性采样长度

为 $\sum_i |E_i| \lfloor w_i / w_{\min} \rfloor$ ，其中 $|E_i|$ 表示 E_i 的长度。

2.3.2 显著性检验器

显著性检验器旨在发现区域抗毁性采样的相对差异，排除随机模拟破坏对抗毁性采样结果造成的波动，发现不同区域抗毁性度量的真实差异。

本节从 2 个角度出发，设计 2 个显著性检验器。第一个角度是破坏结果的整体水平，该角度反映多次随机破坏下区域拓扑的受破坏情况，经典的统计量有平均值、中位数。第二个角度是破坏结果的波动程度，波动较大表明存在某些薄弱区域，经典的统计量有方差。

显著性检验器的输入输出定义如下，输入为 n 个一维向量，每个一维向量代表一个区域抗毁性采样，输出为 $[r_{avg1}, \dots, r_{avgn}, r_{std1}, \dots, r_{stdn}]$ ， r_{avg_i} 表示第 i 个区域在整体水平的显著性评估结果，该值也表示区域 i 在整体情况下的抗毁性排名，2 个差异性不显著的区域有相同的抗毁性排名； r_{std_i} 表示第 i 个区域在波动水平的显著性评估结果，同时也表示区域 i 在波动水平下的抗毁性排名。区域抗毁性度量不是确定值，而是经过显著性检验器处理后的各个区域破坏结果的比较排名值。

显著性检验器实现思路如下。整体水平下的显著性检验器使用 Kruskal-Wallis 检验^[13]，该检验判断中位数是否具有显著性差异，在有显著性差异下，通过 Steel-Dwass 事后检验获得两两比较的结果。根据 2.3.3 节的模拟破坏数据和后续区域破坏数据，通过 K-S 正态检验和方差齐性检验，发现抗毁性破坏结果满足正态分布，而不同区域的破坏结果不满足方差齐性，Kruskal-Wallis 检验方法适用于破坏数据。通过方差齐性 Hartley 检验进行波动水平的比较，看两两方差的差值是否有显著性差异。

2.3.3 采样实验

为了验证随机模拟采样以及使用显著性检验比较的合理性，模拟分别随机生成 30、40 个节点的有向图，链接的数量均为 100，破坏链接数量设置为 1、2、3，对应的单个事件发生概率分别为 p 、 p^2 、 p^3 。所有破坏组合事件的数量为 $166\ 750$ ($C(100,3)+C(100,2)+C(100,1)$)。根据连通性计算所有破坏情况下的破坏影响，即网络中无法连通的节点对的比例，计算每个破坏的结果，用破坏影响乘以发生概率，将所有破坏结果累积获得结果向量。

随机从结果向量中抽取部分样本，比较样本和

结果向量在 2 个角度下是否有显著性差异，图 3 为采样结果，纵坐标格式为“节点数量-边数量-检验方法”，包括 2 个图的 2 种检验方法的结果。图中用符号“+”表明有显著性差异，发现在采样样本数量大于 18 000 时，即采样概率超过 11% 时，结果不存在显著性差异。

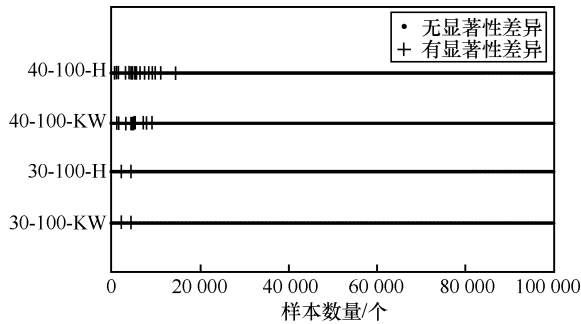


图 3 采样结果

该结果表明，采样率超过 11% 时能反映完整的破坏结果的分布，采样造成的信息损失并不大。

2.3.4 方法评价

已有大量工作针对网络和节点的抗毁性进行分析。其中部分工作根据节点的重要性的网络的脆弱点，提出一些量化指标。一些工作定义度中心性^[14]、介数中心性^[15]来表征节点的中心性。许进^[16]和王梓行^[17]分别提出核与核度、冗余度来表征系统节点的重要性。指标具有可比性，但是仅针对复杂网络，并未加入互联网的路由策略。

此外，大量工作研究复杂网络和互联网的抗毁性。设计不同的破坏角度和方法，包括随机破坏、有针对性破坏、基于互联网的层次结构破坏等。定义衡量破坏的指标主要包括连通度、破坏前后平均路径长度等。文献根据网络在不同破坏模拟下的结果进行分析和讨论。但是此类工作并不能给出具体网络的抗毁性的量化值，因此可比性较差。

目前也有针对区域进行横向比较的工作，但是这些工作均从某些特殊角度进行切入。Leyba 定义 NCP，通过 NCP 评估边界 AS 在区域拓扑的影响力，横向对比 8 个区域，分析 NPC 的变化与差异。Alexander 研究传输运营商在区域层面的影响力，发现哪些区域的流量更不易被其他区域的传输运营商观察和操控流量，具有更高的安全性。

目前尚未调研到有区域整体抗毁性定量比较的方法，本文根据区域抗毁性度量为区域进行排名，排名方法设计的合理性是至关重要的。本文借

鉴经济、社会等领域进行量化排名的已有工作的思想，以量化和比较为目的，并增加互联网资源的权重度量等因素，提出区域抗毁性度量，使排名具有多个角度和更大的灵活性。以下具体说明本文方法考虑的因素及其合理性。

1) 从整体水平和波动水平的角度来共同衡量区域抗毁性。具体原因如下，已通过 K-S 正态检验发现区域的破坏结果数据服从正态分布，而正态分布仅由均值和方差决定，所以整体水平和波动水平一起能捕捉到正态分布数据完整特性。此外，均值和方差在统计学具有重要地位，也应用于各个方面，如 Markowitz 均值-方差模型用来求解最优资产配置的比例，社会收入也会从整体水平和波动水平（基尼系数）来分别讨论，此类工作从不同分析角度出发，评价区域的差异性，对本文工作具有借鉴意义。

2) 用分层概率破坏模拟破坏结果性。大量的文献在研究网络抗毁性时会使用随机破坏的方法，由于随机模拟的无差异性和宏观性，对各个区域的模拟方法相同，使抗毁性比较更具有客观性，抗毁性结果更具有可比性。本文在进行破坏时充分考虑了事件发生的概率，根据事件发生概率模拟破坏事件，符合互联网运行规律。

3) 在区域抗毁性计算中包含多种权重的定义。通过定义区域节点的资源权重（2.1.1 节），衡量不同节点被破坏的差异性。多角度定义的权重支持了模型的可扩展性，提供了多个角度的抗毁性结果，增加了说服力，也让抗毁性结果更加丰富。

4) 聚类并进行分组分析。利用整体水平和波动水平下区域抗毁性的排名结果，对区域进行聚类分组。聚类算法^[18]应用广泛，如构建用户画像、进行恶意流量识别、搜索引擎流量推荐等。本文通过聚类发现组内区域之间的相似性和组间区域之间的差异性。

3 挑战与方法

3.1 计算连通性的高复杂度

计算区域内部的抗毁性度量时，需要知道各个破坏事件下区域内部所有 AS 节点对之间的连通性变化，具有较高的复杂度 $O(n^3)$ ， n 为区域中 AS 的个数。表 1 列出了各区域的 AS 数量和 AS 链接数量，数据去掉了没有客户的 AS 节点。其中美国有 2 119 个 AS 节点，并有 25 个区域 AS 的数量大于 100，降低计算连通性的时间复杂度是很有必要的。

表 1 各区域的 AS 数量和 AS 链接数量(除去没有客户的 AS)

| 区域 | AS 数量/个 | AS 链接数量/个 | 区域 | AS 数量/个 | AS 链接数量/个 |
|----|---------|-----------|----|---------|-----------|
| US | 2 119 | 21 680 | NO | 62 | 236 |
| BR | 807 | 13 134 | IN | 77 | 232 |
| GB | 457 | 8 608 | KR | 61 | 213 |
| DE | 477 | 7 922 | TH | 97 | 202 |
| RU | 1 087 | 7 920 | LU | 40 | 170 |
| NL | 346 | 6 392 | MY | 56 | 163 |
| FR | 345 | 4 898 | BG | 71 | 149 |
| ZA | 309 | 3 776 | CN | 67 | 148 |
| IT | 237 | 3 603 | DK | 50 | 138 |
| CH | 172 | 2 710 | MX | 40 | 113 |
| AU | 317 | 2 187 | AR | 46 | 101 |
| PL | 496 | 1 576 | PH | 45 | 95 |
| CA | 273 | 1 573 | BD | 40 | 81 |
| JP | 255 | 1 200 | CO | 25 | 79 |
| UA | 401 | 1 194 | TR | 34 | 73 |
| SG | 108 | 1 027 | FI | 32 | 72 |
| AT | 125 | 891 | SK | 36 | 68 |
| RO | 114 | 868 | NG | 29 | 63 |
| SE | 144 | 765 | KH | 24 | 60 |
| ES | 122 | 586 | CL | 23 | 49 |
| CZ | 142 | 565 | IR | 34 | 48 |
| IE | 63 | 311 | LV | 31 | 47 |
| NZ | 97 | 309 | PT | 21 | 46 |
| ID | 112 | 242 | HU | 28 | 46 |

本文通过数据预处理来降低计算复杂度，将区域的抗毁性采样转化为区域中 AS 的抗毁性采样。

已有工作^[19]提出了路由树的构建方法，以 AS 节点 AS_{base} 为基础，路由树通过 3 个步骤的广度优先算法，构建出具有路由策略的连通关系树，后文称为 AS_{base} 的路由树。网络中可与 AS_{base} 进行通信的 AS 在且仅在 AS_{base} 的路由树中。计算路由树破坏的连通性复杂度最坏情况下为 $O(m+n)$ ， m 为链接数量， n 为 AS 数量。本文首先构建各个节点的路由树，从计算节点的两两连通性的 $O(n^3)$ 转变为各个节点路由树内与其他节点的连通性，复杂度为 $O(n \times (m+n))$ 。

考虑到各个 AS 的路由影响力不同，只选取重要 AS 来构造路由树并模拟破坏，将复杂度降为 $O(m+n)$ ，得到更加轻量级的计算模型。

3.2 测量数据的缺失

基于探测点收集的 AS 路径数据，很容易缺失关键链路的信息。区域之间边界 AS 的链接对评估区域间通信的抗毁性至关重要。

本文通过发现隐藏链接之间的内在相似性进行缺失链接推断。文献[20]发现隐藏链接的存在，并发现部分隐藏链接的特点。对于只有一个探测点被观察到的链接，将其称为单例链接。如果探测点发生更改，则可能观察不到该链接，造成此类链接的丢失。比如某个观测点观察到路由路径为 $AS_A - AS_{B1} - \dots - AS_{Bn} - AS_C (n \geq 1)$ ，但是另一观测点观察到 AS_A 和 AS_C 也具有直接相连的链接，称 $AS_A - AS_{B1} - \dots - AS_{Bn} - AS_C (n \geq 1)$ 为单例链接 $\langle AS_A, AS_C \rangle$ 的绕行路径，用符号 L_{AC} 表示该链接。此外该论文观察到在所有绕行链接中，连续两跳路径 ($n=1$) 是某个单例链接绕行路径的概率为 80.8%，为大多数情况。

考虑边界 AS 之间的链接，对 $n=1$ 时绕道而行的单例链接进行预测。为了生成训练数据的负样本，当 A 距离观测点的跳数小于或等于 1 时，如果并未观测到链接，则表示不存在该链接，称为负样本。而实际观测到的单例链接则称为正样本。

针对边界链接的特点，表 2 列出了输入特征。本文分别使用 K 近邻回归器、回归树、提升树、随机森林、极端森林、XGB 进行分类预测，通过交叉验证发现，所有算法的准确率均高于 91%，其中极端森林效果最好，准确率达 98%。

表 2 边界 AS 链接预测特征分类及描述

| 分类 | 特征 |
|------------|--|
| AS 关系 | L_{AC} 、 L_{AB} 和 L_{BC} 的 AS 关系 |
| 相关性 | AS_A 和 AS_C 是否相同 |
| | AS_A 和 AS_C 的 ISP 是否相同 |
| | AS_A 和 AS_C 的 ISP 所属区域是否相同 |
| 与边界 AS 的关系 | 以 AS_A 为边界 AS 的区域数量 |
| | L_{AC} 是其他区域边界 AS 链接数量 |
| 观测点位置 | 观测点是否在本区域 |
| 路由变化 | 观测点观测到 AS_A 路由变化数量 |
| | 观测点观测到的路由变化数量 |

本文通过该算法预测到 261 条隐含边界 AS 链接。

3.3 地理定位的准确性

3.3.1 IP 地理定位

IP 定位数据库在区域精度上准确度相对较精确，本文基于 IP 定位数据库实现 IP 的地理定位。为了提高 IP 定位结果的准确率，使用 Team Cymru、Maxmind、IP2location、RIPE IPmap 这 4 个数据库。这 4 个数据库用来对某个 IP 地址查询，可能会发生

错误结果的数量大于正确结果的数量情况，例如 3 个一致的错误结果与一个正确结果。基于合理的假设，认为该现象的发生是少数，设计基于投票的方法将 4 个数据库的结果进行融合。基于投票的方法步骤如下。

步骤 1 计算 4 个数据库的一致性百分比

- 1) 对于每个数据库 D_i ，对每个 IP_i 获得定位结果 C_i 。
- 2) 用其他 3 个数据库对 IP_i 进行地理定位，确定 4 个地理定位结果的多数，并检查 4 个数据库的定位结果是否与多数一致，即是否存在一致性。
- 3) 计算一致性出现的比例，获得数据库 D_i 的一致性百分比 DCR_i 。

步骤 2 根据一致性百分比进行定位结果的投票

- 1) 对于每个 IP_i ，4 个数据库结果分别为 C_1 、 C_2 、 C_3 、 C_4 。
- 2) 每个数据库 D_i 对自己的结果进行投票，投票分数为 DCR_i 。
- 3) 获得投票总分数最多的结果即 IP_i 的定位结果。

3.3.2 AS 地理定位

目前，没有直接方法进行 AS 的地理定位。AS 路径进行地理映射的最大难点在于存在跨越数个区域的 AS，难以对其进行地理定位。虽然地区性互联网注册机构 (RIS, regional internet registry) 在为每个区域分配 AS 号时登记了每一个 AS 所属的区域，但是登记的只是每个 AS 法律意义上的所属区域，实际上的 AS 路径可能实际经过了其他的区域。结合 IP 定位数据库和 Traceroute 数据，本文提出 AS 地理定位的快速方案。

本文处理思路如下。首先获得 AS 在各个区域的分布，IP 定位数据库中将 IP 地址定位到 AS 和区域，以 IP 为媒介，能够获得 AS 和区域的对应关系。其次，对于位于单个区域的 AS，可以直接实现区域定位。计算发现，仅有 6.8% 的 AS 位于多个区域，对于这部分 AS，结合该 AS 的上下文环境 (AS 路径中该 AS 的前后 AS) 进行针对性计算。

借鉴 Karlin 等^[21]提出的区域路径推测算法，本文设计了轻量级的匹配方法，对位于多个区域的 AS 进行地理定位。Traceroute 主动测量能体现自治域内路由的情况，首先将 Traceroute 数据中 IP 路径定位到 AS 和区域，获得 AS 路径和区域路径的对应关系数据库。特别地，对于某个链接

$\langle AS_x, AS_x \rangle$ ，其是位于不同区域的相同 AS 之间的链接，在 AS 路径中合并为一个 AS，这个 AS 的地理定位是 2 个区域。图 4 描述了位于多个区域的 AS 地理定位的 4 个步骤，具体描述如下。

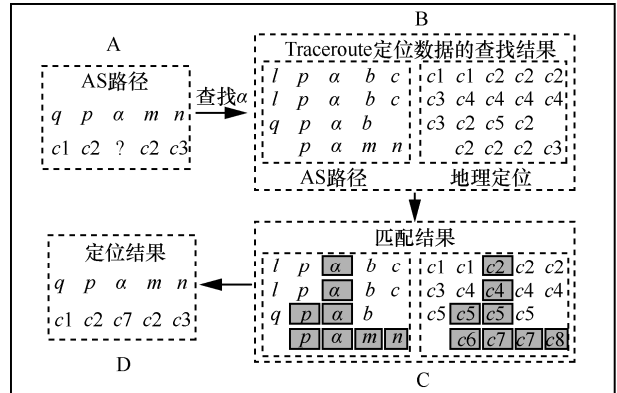


图 4 位于多个区域 AS 的地理定位步骤

- 1) 对于 AS 路径 P 中需要地理定位的 AS_α 。
- 2) 提取在 Traceroute 数据中出现 AS_α 的 AS 路径 Q 。
- 3) 对于每个 AS 路径 Q_i ，将 Q_i 和 P 进行匹配，具体为，以 AS_α 为起点，向前向后获得最长匹配子串 M_i 。
- 4) 最长的子串 M_x 中 AS_α 的定位即 AS 路径 P 中 AS_α 的定位。

4 抗毁性计算

本节首先介绍数据处理，目标为构建具有区域定位标签的全球 AS 拓扑，之后分别介绍区域内部和外部的抗毁性度量的计算细节，抗毁性计算的方法已在 2.3 节介绍。

4.1 数据处理

数据为从 RIPE Atlas 平台收集到的 Traceroute 数据以及从 Routeview 平台收集到的 BGP 数据，二者均为 2020 年 3 月平台的全部数据。此外，由于探测点的局限，预测未发现的边界 AS 链路作为拓扑图的补充。

在数据处理中，需要构建具有区域标签的 AS 级拓扑图。AS 拓扑图中，以 AS 为节点，AS 间建立付费关系 P2C (provider to customer) 或对等关系 P2P (provider to provider) 为边。

这一部分主要包括 2 个步骤：第一，需要分别把 Traceroute 和 BGP 数据中的 IP、AS 定位到具体的 AS 和区域，形成拓扑图中具有区域标签的 AS

节点；第二，为了在后续分析中考虑加入路由策略的限制，需要给每个 AS 链接打上关系标签，学术界普遍将链接的关系分类为 P2C 和 P2P。

学术界有很多推测 AS 关系的工作^[22-24]，并发布了对应的数据集。利用已有的 4 个 AS 关系数据集 (AS Rank、Problink、Toposcope、在 Toposcope 中加入发现的隐藏链接后得到的数据集 h-Toposcope)。

4.2 区域内部抗毁性计算

区域内部关注区域内部拓扑中 AS 之间的连通性。在计算 AS 破坏事件度量时，关注该 AS 与其余 AS 节点的连通。下面具体描述在区域内部，AS 节点资源的定义，以及区域内需要计算抗毁性采样的 AS 的选择。

4.2.1 节点资源定义

从 3 个角度定义 AS 节点 i 的资源权重 r_i 。

- 1) 从连通性角度， r_i 恒为 1。
- 2) 从用户影响力角度， r_i 为 AS _{i} 的用户数量。
- 3) 从域名影响力角度， r_i 为 AS _{i} 的域名重要性影响。

其中，连通性角度默认各个 AS 的资源权重相同，关注于连通的 AS 数量。考虑到网络中各个 AS 的用户数量并不相同，用户影响力角度计算 AS 在全区域用户数量。文献[25]使用前缀顶级列表，将基于域的顶级列表聚合到网络前缀中，获得各个 AS 基于域名重要性的影响度量。使用该工作的结果作为域名重要性影响。

4.2.2 重要 AS 集合的选择

根据区域节点 AS cone 值选择重要 AS，AS cone 定义为其直接和间接客户 AS 的数量。由于只希望选择对本区域抗毁性表征力较强的节点 AS 集合，这里只计算位于本区域的直接和间接客户的 AS 数量。

希望尽可能捕捉到本区域的整体情况，选择方式如下，首先选择 AS cone 排名前 5 的 AS，之后按照 AS cone 排名逐步添加 AS，直到所有 AS 的直接和间接客户覆盖区域 90% 的 AS。

4.3 区域外部抗毁性计算

区域外部抗毁性关注本区域和其他区域的连通性。下面具体描述在区域外部中，区域节点资源的定义，以及需要计算抗毁性采样的 AS 的选择。

4.3.1 节点资源定义

区域外部抗毁性关注本区域和其他区域的连通性，对于区域 i ，从 3 个角度定义其资源权重 r_i 。

- 1) 从连通性角度， r_i 恒为 1。

2) 从区域经济影响力角度， r_i 为区域 i 人均 GDP 的 Zipf 权重。

3) 从域名影响力角度， r_i 为区域 i 的域名重要性影响。

其中，连通性角度默认各个区域的资源权重相同，关注于连通的区域数量。从区域经济实力出发，用区域人均 GDP 度量区域资源大小。已知在互联网中的大量事件流行遵循 Zipf 分布^[26-27]，为了防止区域间人均 GDP 数值差距过大，用 Zipf 分布对人均 GDP 数据重新处理，得到的 Zipf 权重为区域的人均 GDP 权重 r_i 。与 AS 资源权重类似，从域名角度统计区域内部所有 AS 的资源权重作为 r_i 。

4.3.2 重要 AS 集合的选择

选择区域的所有边界 AS 作为重要节点集合，边界 AS 是本区域和其他区域通信的“出入口”，边界 AS 的抗毁性采样结果能全面反映本区域和其他区域通信的情况。

5 实验结果与分析

本节展示抗毁性评价的结果，由于探测点地理分布的差异，有部分区域的测量数据很少。根据测量数据，去除 AS 链接数量太少的区域。本文研究 48 个区域的抗毁性，其中欧洲区域 24 个，亚洲区域 14 个，非洲区域 2 个，北美洲区域 3 个，南美洲区域 3 个，大洋洲区域 2 个。

5.1 区域内部抗毁性

实验具体设计如下，使用 4.1 节提到的 4 个 AS 关系数据集和算法，使用 4.2 节提到的 3 种节点资源权重。观察区域内部抗毁性结果受 AS 关系和 AS 资源权重的影响。模拟破坏节点数量为 1~6，破坏后该节点所有相连的链接被断开。

图 5(a)和图 5(b)分别描述整体水平和波动水平下，各个区域内部抗毁性度量的排名值。类别 1~4 分别表示 AS Rank、Problink、Toposcope、h-Toposcope 数据，b、u、d 分别表示连通性、用户数量、域名重要性。将区域抗毁性度量结果用灰度图显示，区域抗毁性度量即相对排名，排名越高，区域抗毁性越好，灰度越低。灰度图的横坐标表示 48 个区域，用区域代码表示，纵坐标表示不同的 AS 关系数据集和资源权重组合，按照第一个组合下区域抗毁性排名对区域排序展示。

首先分析图 5(a)在整体水平下区域抗毁性度量排名。在 12 种组合下，巴西 BR、美国 US、俄罗

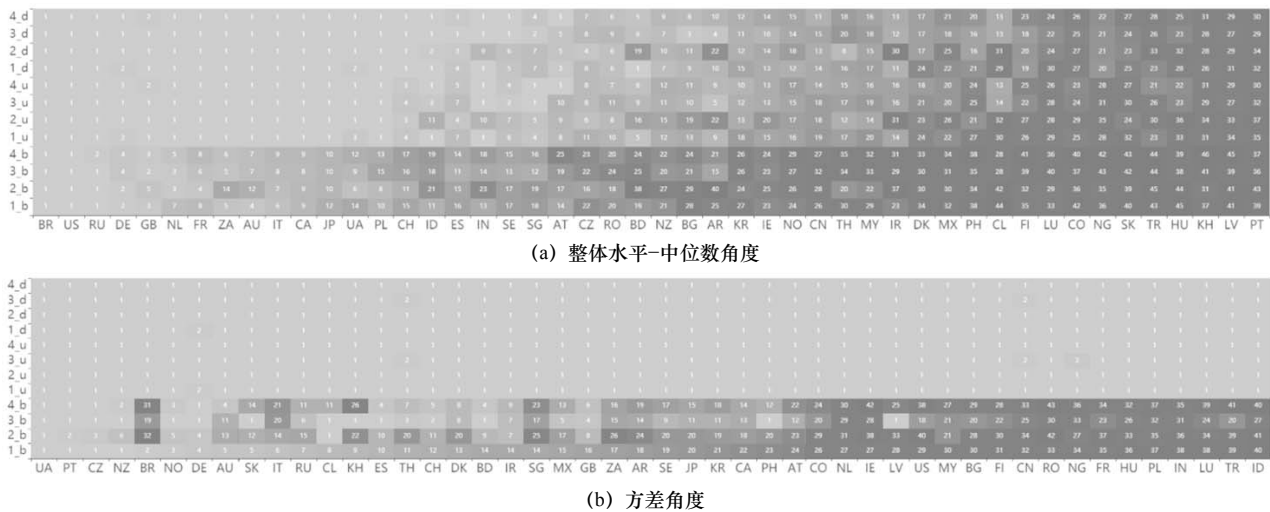


图 5 区域内部抗毁性排名

斯 RU 抗毁性度量排名并列第一。此外，德国 DE、英国 GB、荷兰 NL、法国 FR、南非 ZA、澳大利亚 AU、意大利 IT、加拿大 CA、日本 JP、乌克兰 UA、波兰 PL 抗毁性排名紧随其后，除了以连通性作为资源权重的情况，其余组合中排名大多区域并列第一。可以发现，排名位于前列的区域大多位于欧洲，南非 ZA 在非洲排名第一，日本 JP 在亚洲排名第一，澳大利亚 AU 在大洋洲排名第一。

图 5(b)展示各个区域在波动水平下的区域抗毁性度量的排名。该排序和整体水平下的区域抗毁性度量排名相差很大，中位数和方差没有明显的相关关系，这也间接说明 2 个指标联合评价的合理性。

综合 2 个角度下的区域抗毁性度量，对区域进行聚类，图 6 展示聚类结果。

图 6(a)综合 2 个角度下的区域抗毁性度量排名(12 个组合)，对区域进行聚类，将结果分为六类。纵横坐标分别是整体水平(中位数)和方差角度在 12 个角度下区域抗毁性度量的总和，每个类别按照横坐标由小到大排序。

类别 1 在 2 个角度下都表现较好，特别是俄罗斯 RU、德国 DE、英国 GB、澳大利亚 AU、乌克兰 UA，在整体水平和波动水平下都有较高的抗毁性值。类别 2 和类别 3 有很好的整体水平，但是波动水平下的抗毁性度量稍差。其中美国 US 和巴西 BR 平均水平最好，但是方差角度下抗毁性稍差，属于类别 3。具体观察图 6(b)方差结果，各个区域只有在连通性度量下有明显的差异。这一类的区域

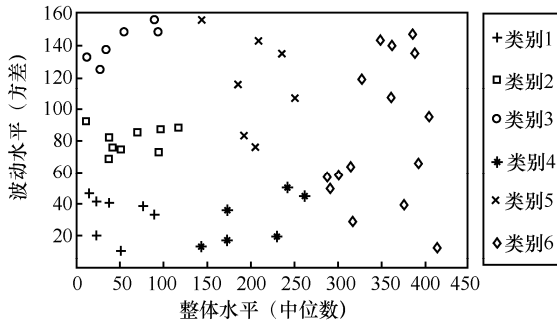
有部分关键链路，破坏后会影响到部分节点的连通，但这些节点重要程度较小，导致对整体网络影响很小。类别 4、类别 5 在 48 个区域中抗毁性处于中等，其中类别 4 在波动水平下抗毁性更好。最后类别 6 表现较差，较之前面类别的区域，受到破坏后对本区域通信的影响较大。

观察到波动水平的结果中，不同资源权重下的结果很不相同，进一步用资源权重划分结果。

图 6(b)展示了资源权重为连通性的聚类结果，划分为七类，该结果与 12 个组合下的聚类结果很相近。主要原因为，在波动水平下，各个区域在其他度量下的排名基本相似，而在整体水平中，各个区域在其他度量下的差异也小于连通性资源权重的差异。

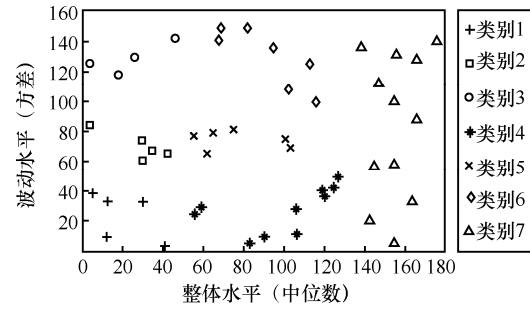
图 6(c)和图 6(d)分别展示了资源权重为用户数量和域名重要性的聚类结果，二者结果相近，且该结果与图 6(b)有较大差异。两张图中，所有区域的方差几乎相同，基本依赖区域的整体水平进行聚类。其中，类别 1 表现最好，该集合数量也很多超过 20 个。之后是类别 2 和类别 3，其抗毁性度量也较接近。前 3 个类别共涉及一半以上的区域，超过一半的区域抗毁性情况相近且均很优秀。之后类别 4 和类别 5 差异增加，抗毁性变差。

从不同度量下的聚类结果可以看出，超过一半的区域有较优秀的抗毁性能力，其中各个区域有一些细微差异而导致波动水平下排名的区别，主要体现在一些区域有部分较脆弱的节点集合，在某些破



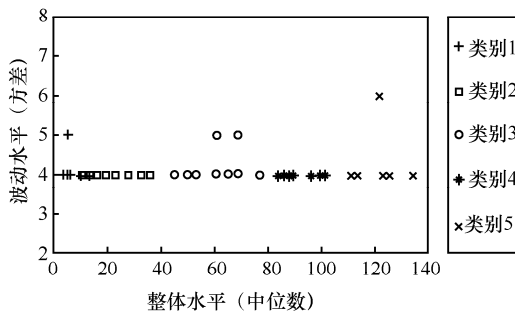
| | | | | | | | | |
|-----|----|----|----|----|----|----|----|----|
| 类别1 | RU | DE | GB | AU | UA | CH | ES | |
| 类别2 | BR | IT | ZA | CA | JP | SE | SG | AT |
| 类别3 | US | NL | FR | PL | ID | IN | | |
| 类别4 | CZ | NZ | BD | NO | TH | IR | | |
| 类别5 | RO | BG | AR | KR | IE | CN | MY | |
| 类别6 | DK | MX | PH | CL | FI | LU | CO | NG |
| | SK | TR | HU | KH | LV | PT | | |

(a) 12个组合下抗毁性度量之和



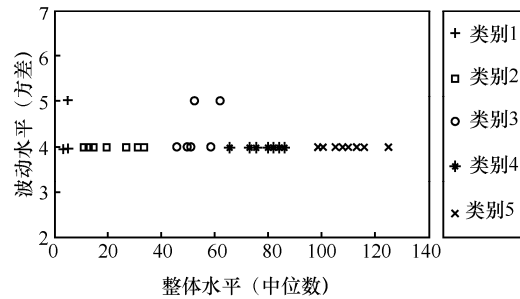
| | | | | | | | |
|-----|----|----|----|----|----|----|----|
| 类别1 | RU | DE | GB | AU | UA | | |
| 类别2 | BR | IT | ZA | CA | JP | | |
| 类别3 | US | NL | FR | PL | | | |
| 类别4 | ES | CH | CZ | NZ | NO | BD | TH |
| 类别5 | IR | DK | MX | | | | |
| 类别6 | SE | SG | AT | AR | KR | | |
| 类别7 | IN | ID | RO | IE | BG | CN | MY |
| | LU | CL | PH | FI | CO | PT | KH |
| | NG | SK | HU | LV | TR | | |

(b) 4个组合下抗毁性度量之和 (连通性)



| | | | | | | | |
|-----|----|----|----|----|----|----|----|
| 类别1 | BR | US | RU | NL | FR | ZA | AU |
| | IT | CA | JP | PL | DE | GB | UA |
| 类别2 | SG | IN | CH | ID | SE | ES | |
| | AT | CZ | BD | RO | | | |
| 类别3 | AR | NZ | BG | KR | IE | TH | NO |
| | CN | MY | IR | | | | |
| 类别4 | DK | MX | CL | PH | FI | TR | CO |
| 类别5 | LU | SK | HU | NG | LV | KH | PT |

(c) 4个组合下抗毁性度量之和 (用户数量)



| | | | | | | | | | | |
|-----|----|----|----|----|----|----|----|----|----|----|
| 类别1 | BR | US | RU | NL | FR | ZA | AU | IT | CA | JP |
| | PL | CH | DE | UA | ID | GB | | | | |
| 类别2 | ES | IN | SE | AT | SG | RO | CZ | BD | BG | NZ |
| 类别3 | AR | KR | IE | CN | NO | TH | | | | |
| 类别4 | IR | MY | PH | DK | FI | NG | CL | MX | | |
| 类别5 | SK | LU | CO | HU | TR | KH | LV | PT | | |

(d) 4个组合下抗毁性度量之和 (域名重要性)

图 6 区域内部抗毁性聚类结果

坏事件下无法连通，但节点集合的总影响力很小。从多个资源权重评价结果是合理且很有意义的，连通性度量更能捕捉网络的差异性，发现网络的轻微变化，但也会导致一些影响结果的“误判”，无法分辨破坏的真实影响力，将用户数量、域名重要性作为资源度量参考有效避免了该问题。

下面，讲述在内部抗毁性结果中的一些发现。

部分区域有抗毁性较差的外围 AS 集群。从图 6(a)发现，AS 关系的变化对区域抗毁性排名影响很小，而不同资源权重下区域的抗毁性具有一定差异。在连通性度量中，各个区域差异性更大。其原因为区域外围有部分 AS 资源权重很小，被破坏后的影响很小。连通性度量认为各个 AS 重要性相同，更加受到该类 AS 的影响。从图 6 中可以发现，

意大利 IT、加拿大 CA、日本 JP、乌克兰 UA、波兰 PL、印尼 ID、印度 IN 这些区域在用户数量、域名重要性度量下排名前列，但是在连通性角度下与巴西 BR 等排名最前的区域差异增大，这些区域都具有抗毁性较高的区域内核 AS 集群，同时区域外围一些资源权重较低的 AS 更易被破坏。

资源权重较高的节点不会被某些破坏事件突然摧毁。图 6(b)中，只有在资源权重为连通性时，各区域才出现明显的差异，在其余的资源权重指标下，各个区域排名极其接近。该数据说明在各个破坏情况下，被破坏后无法通信的节点数量相差比较大，但是总体其在用户数量和域名重要性角度下，相差很小，间接反映了各个区域有较高用户数量和域名重要性的节点很少会因为某些特别的攻击而

受到直接的影响。

所有区域对别的区域管理的 AS 依赖性很小。对于每个区域，图 7 只破坏区域内部由其他区域管理的 AS，旨在发现是否有部分区域内部拓扑大大依赖于其他区域管理的 AS。与图 5(a)相比，3 个资源权重下区域之间的差异减小，在资源权重为用户数量和域名重要度下，大约 30 个区域的排名均处于第一和第二，这些区域中非本区域管理的 AS 对区域核心 AS 集群影响很小。多数区域的排名均提高，其中澳大利亚 AU、南非 ZA、瑞士 CH、印度 IN 在资源权重为连通性时排名大大提高，整体排名也随之得到提高。伊朗 IR、马来西亚 MY 在资源权重为用户数量和域名重要性下的提高很明显，二者在图 5(a)中 3 个度量下的排名均大于 10，但在图 7 中用户数量和域名重要性中排名均上升到前两名。说明 2 个区域虽然整体安全抗毁性较差，但其他区域管理的 AS 在该区域重要性很低，该区域能较好地抵御其他区域管理 AS 的针对破坏。

5.2 区域外部抗毁性

区域外部抗毁性观察区域之间的连通，使用 4.3 节提到的 3 种区域节点资源权重。观察抗毁性结果发现，不同 AS 数据集的结果差异很小，因此图中只展示了使用 AS Rank 数据集的结果。

从 2 个角度出发，图 8(a)描述各个区域外部抗毁性度量即排名值。灰度图横坐标表示 48 个区域，纵坐标 b、g、d 分别表示连通性、经济影响力、域名重

要性度量，var 表示方差角度，无标注表示中位数角度。按照 6 个组合下区域抗毁性排名的均值对区域排序。与区域内部抗毁性结果图 5 相比，各区域的抗毁性差距拉大，很少存在并列排名，各区域的具体排名也有变化。从图 8 中可以发现，不同资源权重下结果差距很小，分析原因是各区域边界 AS 数量有较大差异，所以资源权重对结果影响很小。

图 9 展示区域外部抗毁性在整体水平和波动水平 2 个角度下的聚类结果，由于不同资源权重下结果差异很小，这里只展示了全部组合的聚类结果。从聚类结果可以看到，整体水平和波动水平的结果高度正相关，将各个区域划分为 5 个类别，其中类别 1 和类别 2 区域外部抗毁性较好，美国 US、德国 DE、英国 GB、加拿大 CA 和俄罗斯 RU 属于类别 1，抗毁性最好。不同于区域内部抗毁性结果中大量区域集中在较好的类别，区域外部抗毁性中大量区域集中在较差的类别，其中超过 80% 的区域都在后 3 个类别中。

下面，讲述在外部抗毁性结果中的一些发现。

不存在一些区域大量依靠其他区域管理的边界 AS 服务。考虑到区域外部抗毁性和边界 AS 抗毁性高度相关，但是位于本区域的边界 AS 并不一定由本区域管理，本文只计算位于本区域且由本区域管理的边界 AS，图 8(b)展示该实验的抗毁性结果。和图 8(a)相比较，各个区域排名都很类似，差距不大，所以不存在一些区域大量依靠其他区域管

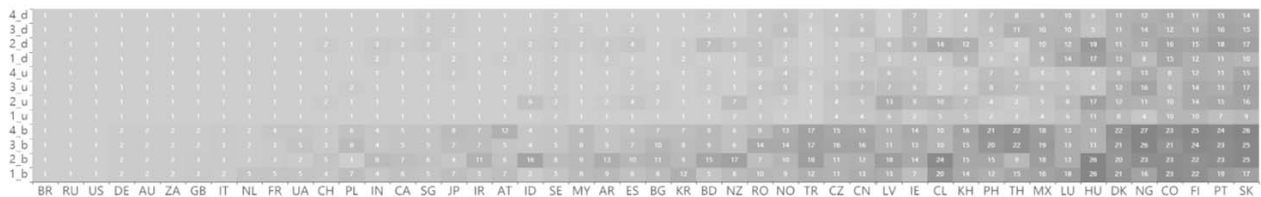
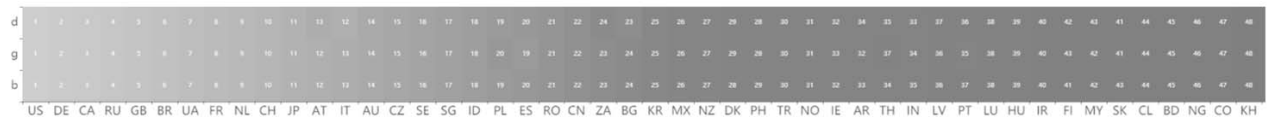


图 7 区域内部抗毁性排名（只破坏其他区域管理的 AS 节点）



(a) 全部边界 AS

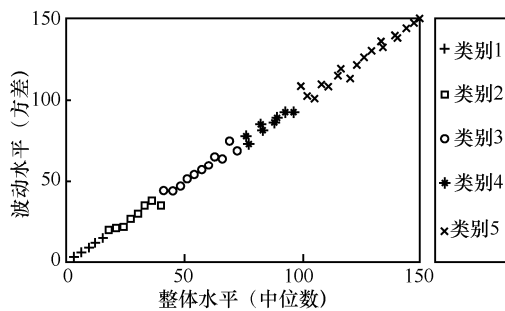


(b) 只有本区域管理的边界 AS

图 8 区域抗毁性外部排名

理的边界 AS 传输稳定的服务。

各个区域边界 AS 数量差异明显。表 3 列出了各个区域边界 AS 的具体数量,用 M_b 表示该区域管理的边界 AS 集合,用 L_b 表示位于该区域的边界 AS 集合。表 3 中数据证明,各个区域边界 AS 数量差距很大,美国 US 有超过 500 的边界 AS,而有 5 个区域数量小于 10。 $M_b \& L_b$ 表示二者交集,即位于本区域且由本区域管理的边界 AS; $M_b - L_b$ 表示二者差集,即位于本区域且不由本区域管理的边界 AS。这两行数据可以得到区域内部边界 AS 由本区域管理的比例,比较各个区域的数据,除了边界 AS 数量较小的区域,约 80%都由本区域管理,其中俄罗斯 RU 的比例特别大,95%的边界 AS 均由自己管理。但是新加坡 SG 有 44%的 AS 由其他区域管理。



| | | | | | | | | | | |
|-----|----|----|----|----|----|----|----|----|----|----|
| 类别1 | US | DE | GB | CA | RU | | | | | |
| 类别2 | FR | NL | BR | UA | CH | JP | AU | AT | | |
| 类别3 | IT | SG | CZ | SE | RO | ES | PL | ID | ZA | |
| 类别4 | BG | CN | DK | NZ | IE | MX | PH | KR | | |
| 类别5 | AR | NO | TR | IN | TH | LU | LV | PT | FI | HU |
| | MY | CL | NG | SK | BD | CO | KH | | | |

图 9 区域外部抗毁性聚类结果

抗毁性排名和区域的边界 AS 数量排名接近。比较抗毁性排名和边界 AS 数量排名,发现二者趋势相同,但也有部分区域不太一致,其中加拿大 CA、日本 JP 边界 AS 数量较少,但是抗毁性排名高,说明该区域具有抗毁性排名较高的边界 AS,也有波兰 PL、保加利亚 BG、印度 IN,边界 AS 数量较多,但是抗毁性排名相对偏低,说明该区域边界 AS 的通信质量拖了后腿。进一步探究原因,将所有区域所有边界 AS 的抗毁性采样输入显著性检验器,得到区域边界 AS 的抗毁性排名。图 10 绘制了 7 个区域的边界 AS 抗毁性度量的排名分布。横坐标为边界 AS 抗毁性排名,纵坐标为对应 AS 的路由影响力占比。从图 10 中可以发现,加拿大 CA 和日本 JP 在第一名的路由影响力比例是最高的。

表 3 区域边界 AS 的数量情况

| 区域 | $ M_b $ | $ M_b - L_b $ | $ M_b \& L_b $ | $ L_b - M_b $ |
|----|---------|---------------|----------------|---------------|
| US | 518 | 188 | 330 | 67 |
| DE | 210 | 38 | 172 | 40 |
| GB | 198 | 52 | 146 | 47 |
| RU | 166 | 19 | 147 | 9 |
| FR | 117 | 23 | 94 | 20 |
| NL | 112 | 29 | 83 | 34 |
| UA | 105 | 3 | 102 | 11 |
| BR | 104 | 3 | 101 | 11 |
| CA | 99 | 6 | 93 | 15 |
| CH | 70 | 6 | 64 | 12 |
| IT | 67 | 1 | 66 | 10 |
| AT | 59 | 12 | 47 | 11 |
| PL | 58 | 6 | 52 | 6 |
| AU | 56 | 8 | 48 | 16 |
| ID | 48 | 3 | 45 | 3 |
| SG | 48 | 10 | 38 | 30 |
| JP | 47 | 4 | 43 | 12 |
| BG | 47 | 4 | 43 | 7 |
| RO | 42 | 11 | 31 | 8 |
| SE | 39 | 6 | 33 | 11 |
| IN | 39 | 5 | 34 | 7 |
| CZ | 38 | 3 | 35 | 10 |
| ZA | 36 | 3 | 33 | 11 |
| ES | 34 | 3 | 31 | 13 |
| CN | 28 | 6 | 22 | 5 |
| NZ | 26 | 3 | 23 | 4 |
| TH | 25 | 0 | 25 | 2 |
| LU | 19 | 3 | 16 | 3 |
| AR | 19 | 3 | 16 | 3 |
| MX | 19 | 0 | 19 | 5 |
| PH | 18 | 0 | 18 | 2 |
| NO | 17 | 4 | 13 | 2 |
| MY | 17 | 1 | 16 | 3 |
| DK | 16 | 1 | 15 | 8 |
| IE | 16 | 0 | 16 | 11 |
| BD | 15 | 0 | 15 | 0 |
| SK | 15 | 1 | 14 | 1 |
| HU | 15 | 3 | 12 | 2 |
| KR | 14 | 0 | 14 | 3 |
| CL | 12 | 1 | 11 | 2 |
| LV | 12 | 0 | 12 | 3 |
| FI | 11 | 2 | 9 | 3 |
| IR | 10 | 0 | 10 | 1 |
| PT | 9 | 0 | 9 | 3 |
| NG | 9 | 2 | 7 | 0 |
| TR | 9 | 1 | 8 | 2 |
| KH | 9 | 0 | 9 | 0 |
| CO | 7 | 2 | 5 | 2 |

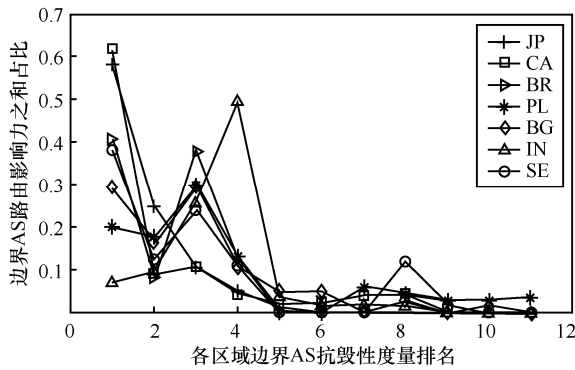


图 10 区域边界 AS 抗毁性排名分布

5.3 区域通信服务影响力与抗毁性的关系

各个区域边界 AS 的抗毁性决定了全球通信的抗毁性，从抗毁性的角度出发，分析各个区域管理的边界 AS 在全球通信服务的重要性，即该区域通信服务的影响力。

思路概括如下。对于每个区域，每次只破坏该区域管理的边界 AS，模拟破坏区域边界 AS 的数量为 1~25，破坏一个边界 AS 即破坏该边界 AS 的全部边界 AS 链接。计算该破坏下全部区域所有边界 AS 的抗毁性采样，汇总得到区域的通信服务影响力。

通信服务影响力和抗毁性排名差异很大。图 11 绘制区域通信服务的影响力排名。横坐标为各个区域，纵坐标为 3 种资源权重。从图 11 中可以发现，约一半的区域影响力都非常小，只破坏这些区域管理的边界 AS，对区域间的通信几乎不造成任何影响。美国 US 和德国 GB 依旧排名靠前，其余与区域抗毁性排名很不一致，如法国 FR、乌克兰 UA、奥地利 AT 在区域抗毁性中都排名靠前，但是服务影响力排名靠后。相反，印尼 ID、捷克 CZ、新西兰 NZ 抗毁性排名较低，区域管理的边界 AS 数量 M_0 较少，但是服务影响力排名位于前列。

进一步分析原因，表 4 列出 7 个区域的边界 AS 抗毁性度量排名前五级的数量。和服务影响力排名一致，美国有很强的影响力，抗毁性排名第一级的边界 AS 数量很多。法国 FR、乌克兰 UA 和奥地利 AT 抗毁性排名前五级的边界 AS 数量只有一个或 2 个，远不如其他区域。印尼 ID、捷克 CZ、新西兰 NZ 抗毁性排名前两级数量很少，但是前五级

的数量很多。法国 FR、乌克兰 UA 和奥地利 AT 的区域通信影响力低的原因在于它们的边界 AS 有很丰富的 AS 链接，且该区域与周围区域也有丰富的 AS 链接，导致破坏该区域管理的边界 AS 对互联网破坏力很小。与之对应的印尼 ID、捷克 CZ、新西兰 NZ 对一些区域有较高的通信影响。

表 4 通信服务影响力下部分区域边界 AS 抗毁性排名数量

| 区域 | 级别 1 | 级别 2 | 级别 3 | 级别 4 | 级别 5 |
|----|------|------|-------|-------|-------|
| US | 285 | 404 | 3 231 | 2 200 | 759 |
| FR | 0 | 0 | 0 | 0 | 2 |
| UA | 0 | 0 | 1 | 0 | 0 |
| AT | 0 | 0 | 0 | 2 | 0 |
| ID | 0 | 5 | 201 | 3 385 | 2 883 |
| CZ | 1 | 26 | 429 | 3 146 | 2 873 |
| NZ | 0 | 9 | 363 | 3 184 | 2 876 |

6 结束语

全球互联网的重要性与日俱增，网络拓扑是互联网通信的基础。本文采集多方数据，建立了较完善的区域粒度的网络拓扑。量化不同区域网络拓扑的抗毁性，定义显著性检验器，从内部拓扑和外部通信 2 个方面对区域抗毁性进行排名并发现区域差异。后续工作将重点考虑发现并定位拓扑中薄弱点，精准给出区域拓扑优化建议。

参考文献：

- [1] KOZOWSK A. Comparative analysis of cyberattacks on Estonia, Georgia and Kyrgyzstan[J]. European Scientific Journal, 2014, 10(7): 237-245.
- [2] OGUN M. Terrorist use of cyberspace and cyber terrorism: New challenges and responses[M]. IOS Press, 2015.
- [3] DAINOTTI A, AMMANN R, ABEN E, et al. Extracting benefit from harm: using malware pollution to analyze the impact of political and geophysical events on the Internet[J]. Computer Communication Review: A Quarterly Publication of the Special Interest Group on Data Communication, 2012, 42(1): 32-39.
- [4] DAINOTTI A. Lessons learned by “measuring” the internet during/after the sandy storm[C]//Invited talk at FCC Workshop on Network Resiliency, 2013.
- [5] ALBERT R, JEONG H, BARABASI A L. Error and attack tolerance of complex networks[J]. Nature, 2000, 406(6794): 378-382.

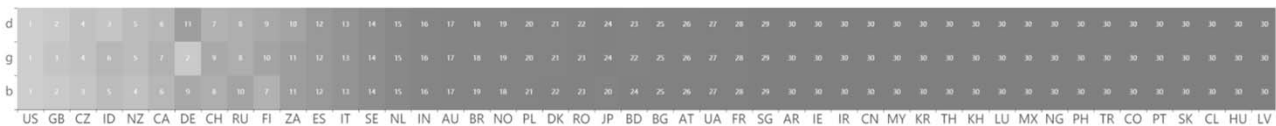


图 11 区域通信服务影响力排名

- [6] BAUMANN A, FABIAN B. How robust is the Internet? – insights from graph analysis[C]//Lecture Notes in Computer Science. Berlin: Springer, 2015: 247-254.
- [7] WU J, ZHANG Y, MAO Z M, et al. Internet routing resilience to failures: analysis and implications[C]//Proceedings of the 2007 ACM CoNEXT Conference on - CoNEXT '07. New York: ACM Press, 2007: 1-12.
- [8] DOLEV D, JAMIN S, MOKRYN O, et al. Internet resiliency to attacks and failures under BGP policy routing[J]. Computer Networks, 2006, 50(16): 3183-3196.
- [9] DOMBROWSKI S, ERMAKOVA T, FABIAN B. Graph-based analysis of cloud connectivity at the internet protocol level[J]. International Journal of Communication Networks and Distributed Systems, 2019, 23(1): 117.
- [10] OMER M, NILCHIANI R, MOSTASHARI A. Measuring the resilience of the global Internet infrastructure system[C]//Proceedings of 2009 3rd Annual IEEE Systems Conference. Piscataway: IEEE Press, 2009: 156-162.
- [11] LEYBA K G, EDWARDS B, FREEMAN C, et al. Borders and gateways: measuring and analyzing national as chokepoints[C]// Proceedings of Proceedings of the Conference on Computing & Sustainable Societies - COMPASS 19. New York: ACM Press, 2019: 184-194.
- [12] FONTUGNE R, SHAH A, ABEN E. The (thin) bridges of AS connectivity: measuring dependency using AS hegemony[C]//Passive and Active Measurement. Berlin: Springer, 2018: 216-227.
- [13] 吴喜之. 统计学: 从数据到结论[M]. 北京: 中国统计出版社, 2009. WU X Z. Statistics: from data to conclusion[M]. Beijing: China Statistics Press, 2009.
- [14] WOLFE A W. Social network analysis: methods and applications[J]. American Ethnologist, 1997, 24(1): 219-220.
- [15] BONACICH P. Factoring and weighting approaches to status scores and clique identification[J]. The Journal of Mathematical Sociology, 1972, 2(1): 113-120.
- [16] 许进. 系统核与核度理论及其应用[M]. 西安: 西安电子科技大学出版社, 1994. XU J. System kernel and kernel degree theory and its application[M]. Xi'an: Xidian University Press, 1994.
- [17] 王梓行, 姜大立, 漆磊, 等. 基于冗余度的复杂网络抗毁性及节点重要度评估模型[J]. 复杂系统与复杂性科学, 2020, 17(3): 78-85. WANG Z H, JIANG D L, QI L, et al. Complex network invulnerability and node importance evaluation model based on redundancy[J]. Complex Systems and Complexity Science, 2020, 17(3): 78-85.
- [18] 章永来, 周耀鉴. 聚类算法综述[J]. 计算机应用, 2019, 39(7): 1869-1882. ZHANG Y L, ZHOU Y J. Review of clustering algorithms[J]. Journal of Computer Applications, 2019, 39(7): 1869-1882.
- [19] GILL P, SCHAPIRA M, GOLDBERG S. Modeling on quicksand[J]. ACM SIGCOMM Computer Communication Review, 2012, 42(1): 40-46.
- [20] JIN Z T, SHI X G, YANG Y, et al. TopoScope: recover AS relationships from fragmentary observations[C]//Proceedings of the ACM Internet Measurement Conference. New York: ACM Press, 2020: 266-280.
- [21] KARLIN J, FORREST S, REXFORD J. Nation-state routing: censorship, wiretapping, and BGP[R]. 2009.
- [22] GAO L X. On inferring autonomous system relationships in the Internet[J]. IEEE/ACM Transactions on Networking, 2001, 9(6): 733-745.
- [23] DIMITROPOULOS X, KRIOUKOV D, HUFFAKER B, et al. Inferring AS relationships: dead end or lively beginning? [C]//Experimental and Efficient Algorithms. Berlin: Springer, 2005: 113-125.
- [24] JIN Y, SCOTT C, DHAMDHERE A, et al. Stable and practical AS relationship inference with problink[C]//16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19). Berkeley: USENIX Association, 2019: 581-598.
- [25] NAAB J, SATTLER P, JELTEN J, et al. Prefix top lists: gaining insights with prefixes from domain-based top lists on DNS deployment[C]//Proceedings of Proceedings of the Internet Measurement Conference. New York: ACM Press, 2019: 351-357.
- [26] LI W T. Zipf's Law everywhere[J]. Glottometrics, 2002, 5: 14-21.
- [27] KRASHAKOV S A, TESLYUK A B, SHCHUR L N. On the universality of rank distributions of website popularity[J]. Computer Networks, 2006, 50(11): 1769-1780.

[作者简介]



安常青 (1970-), 女, 江苏徐州人, 清华大学副研究员, 主要研究方向为计算机网络体系结构、互联网路由、网络协议、网络测量。

刘玉家 (1997-), 女, 山西忻州人, 清华大学硕士生, 主要研究方向为网络测量。

王会 (1977-), 女, 河南南阳人, 博士, 清华大学副教授、博士生导师, 主要研究方向为互联网体系结构、互联网测量、大数据相关技术。

郑志延 (1979-), 男, 北京人, 清华大学工程师, 主要研究方向为互联网路由、网络管理、未来网络试验设施。

喻涛 (1981-), 男, 湖北荆州人, 博士, 清华大学助理研究员, 主要研究方向为网络体系结构、网络空间安全、未来互联网试验床、SDN/人工智能/大数据应用。

王继龙 (1973-), 男, 黑龙江大兴安岭人, 博士, 清华大学教授, 主要研究方向为网络空间测绘、位置网、未来网络试验设施等。