

网络空间测绘技术研究

赵帆^{1,2}, 罗向阳^{1,2}, 刘粉林^{1,2}

(1. 解放军信息工程大学网络空间安全学院, 河南 郑州 450001;

2. 解放军信息工程大学数学工程与先进计算国家重点实验室, 河南 郑州 450001)

摘要: 网络空间测绘技术是一种网络共性基础与应用技术, 对网络空间测绘相关技术的研究进展进行了综述和讨论。首先, 阐述了网络空间测绘的内涵, 给出了网络空间测绘一般技术框架; 然后, 从探测层、映射层和绘制层分别阐述了网络空间测绘相关技术的研究进展; 最后, 指出了网络空间测绘亟待解决的一些关键问题。

关键词: 网络空间测绘; 探测; 映射; 绘制; 地理空间; 社会空间

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2016.00097

Research on cyberspace surveying and mapping technology

ZHAO Fan^{1,2}, LUO Xiang-yang^{1,2}, LIU Fen-lin^{1,2}

(1. School of Cyberspace Security, Zhengzhou Science and Technology Institute, Zhengzhou 450001, China;

2. State Key Laboratory of Mathematical Engineering and Advanced Computing,
Zhengzhou Science and Technology Institute, Zhengzhou 450001, China)

Abstract: Cyberspace surveying and mapping technology is a network common foundation and application technology. The research progress of cyberspace surveying and mapping technology was reviewed and discussed. Firstly, the connotation and principle framework of cyberspace surveying and mapping technology was described. Secondly, relevant technologies were introduced in probing layer, mapping layer and drawing layer. Finally, some open problems in this field were discussed, and some interesting directions that may be worth researching in the future were indicated.

Key words: cyberspace surveying and mapping, probing, mapping, drawing, geographic space, social space

1 引言

随着计算机技术的发展, 网络已经成为国家政治、经济、军事等社会系统存在和发展的基础, 网络空间已经成为人类生产生活的“第二类生存空间”, 网络一旦受损, 几乎所有的社会系统都将无法运行或发挥作用。因此, 亟需对网络空间实施高效的管理、合理的资源分配以及有效

的安全监测和防护。

地图是描绘地理空间信息的重要载体, 自古以来就是运筹帷幄、指挥作战的重要工具, 是指挥员的“眼睛”“行军的无声向导”“协同作战的共同语言”。随着互联网的发展, 地图的应用越来越广, 已成为面向位置的服务(LBS)的核心基础。然而, 网络空间尚缺乏类似地理空间地图的、可全面描述和展示网络空间信息的“网络地图”。

收稿日期: 2016-06-17; 修回日期: 2016-08-06。通信作者: 罗向阳, xiangyangluo@126.com

基金项目: 国家自然科学基金资助项目(No.61379151, No.61272489, No.61302159, No.61401512, No.61572052); 河南省科技创新杰出青年基金资助项目(No. 144100510001)

Foundation Items: The National Natural Science Foundation of China (No.61379151, No.61272489, No.61302159, No.61401512, No.61572052), The Technology Innovation Excellent Youth Foundation of Henan Province (No. 144100510001)

类似于地理空间测绘，构建网络空间地图的技术称为“网络空间测绘”。

网络空间测绘技术通过网络探测、采集、处理、分析等方式，获得网络空间基础设施、用户及服务网络资源在网络空间、地理空间以及社会空间的相关属性，并将这些属性以逻辑图和地理信息图的形式绘制出来，从而直观实时地反映出当前网络空间资源各个属性的状态、发展趋势等。

当前，与网络空间测绘技术相关的研究工作已经开展，如美国国防部国防高级研究计划局于 2012 年 9 月启动的网络战发展项目“X 计划”^[1]，目的是生成网络空间作战态势图、制定作战方案、实施网络作战行动等；美国国家安全和英国国家通信总局（GCHQ）联合开展的研究项目“藏宝图（treasure map）”计划，聚焦于逻辑层捕获路由及自治系统的数据，试图绘制出一张“近乎实时的、交互式的全球互联网地图”^[2]。上述研究的主要技术均属网络空间测绘研究范畴，可见，网络空间测绘技术在国家安全和军事领域具有十分广泛的应用前景。

本文首先阐述网络空间测绘概念的基本内涵，然后从探测层、映射层和绘制层 3 个层次分别介绍网络空间测绘相关的关键技术及研究现状，最后指出网络空间测绘技术亟待解决的一些关键问题。

2 网络空间测绘基本概念

2.1 网络空间测绘技术的定义

目前，尚未见正式的公开文献对网络空间测绘技术进行严格定义。本文从狭义和广义 2 个角度分别阐述网络空间测绘技术的基本内涵。

从狭义上讲，网络空间测绘技术主要指在互联网环境下，利用网络探测、采集或挖掘等技术，获取网络设备等实体资源、用户和服务等虚拟资源的网络属性，通过设计有效的定位算法和关联分析方法，将实体资源映射到地理空间，将虚拟资源映射到社会空间，并将探测结果和映射结果绘制出来。

从广义上讲，网络空间测绘的研究范围包括互联网、电信网、工业控制网等各种类型的网络，探测对象除互联网资源外，还包括其他各种网络

上的资源。

本文讨论的主要是狭义的网络空间测绘技术。

2.2 网络空间测绘主要研究内容

网络空间测绘技术研究对象包括实体资源和虚拟资源两类。实体资源根据设备用途可分为网络基础设施和接入设备，也可根据有无 IP 分为有 IP 化的实体网元和无 IP 化的基础资源；虚拟资源包括网络虚拟人物，网络虚拟社区，以及文本信息内容、音视频、网站等网络服务等。

网络空间测绘技术的目标是实现对来源众多、类型各异的互联网资源的全面测绘，涉及的技术较多。本文从探测层、映射层和绘制层 3 个层次分别给出网络空间测绘技术的相关研究内容，总体框架如图 1 所示。

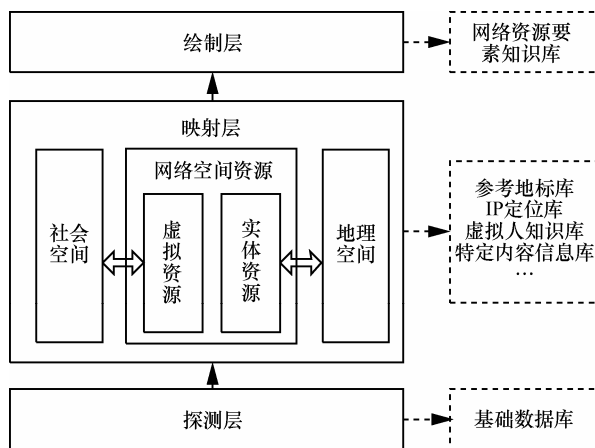


图 1 网络空间测绘技术总体框架

其中，探测层是网络空间测绘技术的基础层，为映射层提供探测基础数据，主要研究实体资源的拓扑探测技术，组件识别技术和对文本信息内容、音视频、网站等的探测分析技术，可简要归纳为探测通道与平台技术、多种探测技术和探测结果分析技术；映射层是网络空间测绘技术的核心层，主要研究实体资源向地理空间映射技术和网络虚拟人物、虚拟社区等虚拟资源向社会空间映射技术，并将映射结果提供给绘制层；绘制层将探测结果和映射结果可视化，主要研究逻辑图绘制技术和地理信息图绘制技术。

下面分别介绍网络空间测绘中探测层、映射层和绘制层主要相关技术基本原理及研究进展。

3 网络空间测绘探测层关键技术与研究进展

3.1 探测层技术框架

探测层是网络空间测绘技术的基础层,主要作用是基于统一的高效探测通道和平台,利用针对特定对象的探测技术,分别获取网络实体资源和虚拟资源的相关网络属性,并对探测结果进行分析。图2为探测层的总体技术框架。

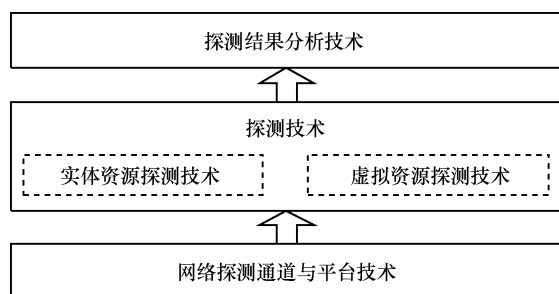


图2 探测层技术框架

如图2所示,探测层研究的主要内容包括网络探测通道与平台技术、探测技术和探测结果分析技术。其中,根据探测对象的不同,探测技术分为实体资源探测技术和虚拟资源探测技术两类。下面分别简要介绍这3种技术。

3.2 探测通道与平台技术

网络空间资源多样、规模庞大,为了能够高效、迅速地对网络空间资源发起探测,且不影响网络的正常运行,需要构建统一的网络探测通道与平台。探测通道与平台技术通过构建统一的分布式网络探测平台,实现广泛分布式部署的探测终端统一化管理和高效持续探测,为发起可靠探测提供探测环境和技术保障。目前,分布式探测系统已成为研究人员了解网络状态的首选工具。比较有代表性的分布式探测系统有塞浦路斯大学的IPMicra^[3]、比尔肯大学的SE4SEE^[4]、佐治亚理工学院的Apoidea^[5]等。Spring等^[6]的RocketFuel项目首次全面探测了大型ISP拓扑。Katz-Bassett^[7]实现了reverse-traceroute,利用时间戳和记录路由选项来识别反向路径。在扫描工具方面有业内常用的功能强大的Nmap^[8],也有最近出现的以快著称、45 min扫描整个IPv4空间的Zmap^[9]。

3.3 探测技术

根据探测的对象不同,网络空间探测技术可分为实体资源探测技术和虚拟资源探测技术两类,根据探测的方式不同,网络空间探测技术还可分为主动探测和被动探测两类,本文主要根据前者分类进行介绍。

3.3.1 实体资源探测技术

实体资源探测技术的探测目标包括网络基础设施和各种接入设备,涉及的技术主要包括网络拓扑发现技术和网络设备组件识别技术等。

网络层拓扑发现一般可分为4个层次:IP接口级拓扑发现、路由器级拓扑发现、入网点(PoP, point of presence)级拓扑发现和AS(自治系统)级拓扑发现。对网络空间测绘而言,通常更关注前3个层次。在IP接口级拓扑测量方面,现有主要方法包括基于简单网络管理协议(SNMP, simple network management protocol)^[10,11]、基于网际组管理协议(IGMP, Internet group management protocol)^[12~14]、基于Traceroute工具等拓扑测量方法。在路由器级网络拓扑测量方面,研究的重点是解决路由器别名问题,即别名解析^[15]。近年来,研究者们主要利用IP时间戳选项来解决别名归并的问题。Sherry等^[16]提出了基于IP时间戳选项的别名判别方法,Marchetta等^[17]提出了基于时间戳选项和UDP的Pythia算法来用于别名解析。在PoP级网络拓扑测量方面,首次尝试探索PoP级网络拓扑的是Spring等^[18],提出了基于DNS解析的PoP提取算法,通过对大量探测结果进行分析,以网络服务提供商(ISP)的命名惯例提取合并不同的PoP。Madhyastha等^[19]提出iPlane方法,利用Mercator方法对各个IP进行别名归并,并通过Rocketfuel和Sarangworld这2个数据源,将DNS域名指定到尽可能多的网络接口。Feldman等^[20]提出的PoP提取算法通过延迟进行初步划分,接着采取一种结构性的方法进一步划分,在初步划分的结果中寻找具有特定结构的二分子图,进而完成PoP的划分提取。Shavitt等^[21,22]提出了基于IP地理位置数据库和时延测量提高PoP级的网络拓扑定位准确性的算法。经过多年研究,网络拓扑探测与分析技术研究已取得很大进展。然而,如何得到完整准确的网络拓扑结构

仍存在一些亟待研究的问题,探测路径中的匿名路由器、路由器的别名归并等仍是研究的难点。对网络空间测绘技术而言,探测分析区域性的网络,得到探测目标所在区域的网络拓扑结构更具针对性,也是需解决的关键问题之一。

网络设备组件的探测识别技术能够为网络空间测绘提供更详细的设备信息,如设备类型、操作系统等。该技术通常通过设备在通信过程中携带的某种信息对其进行认证和辨识,实现的方式主要包括主动和被动2种。其中,主动式主要是从探测源将构造的信息发送给待识别的目标设备,目的是通过触发某些期望的设备回应来对设备进行辨识;被动式不主动向目标发送信息,而是根据设备向外发送请求时,从其中携带的信息来分析判断。被动的识别方式对应用环境的要求较低,对网络状态影响较小,但其能够提供的信息量和可靠性不如主动式。目前,针对设备识别的研究通常是被动式的。Kohno等在文献[23]中提出了根据设备时钟的偏移值对远程网络设备进行识别的方法,Fink^[24]对该方法进行了改进,引入线性回归的统计方法来判断时钟偏移,并给出了需要某种程度的准确度时样本数量的计算公式,使这种设备识别的准确度可控。

3.3.2 虚拟资源探测技术

在网络空间测绘技术的相关研究中,开展虚拟资源探测,涉及的现有技术主要包括特定信息内容快速探测和话题发现技术、音视频内容探测技术、网站自动探测技术等。目前,对上述技术的研究已取得了一定进展。

在文本资源探测方面,主要包括内容分析、关联分析、社会信息网络挖掘等问题。向量空间模型(VSM, vector space model)、潜在语义模型(LSI, latent semantic indexing)等都试图在语法甚至语义层面给2个文本一个相似性度量^[25-27]。在特定信息内容快速探测和话题发现方面,文献[28]提出利用概念性动态隐含狄利克雷(CDLDA, conceptual dynamic latent dirichlet allocation)模型对文本中的动词和名词的比例分析,来对会话内容中的话题进行检测和追踪;文献[29]提出利用增量聚类框架,并应用一系列内容和时间相关的特点来探测新的热门话题。特定音频内容快速探

测分析包括多语言识别和固定音频检索。2006年,Hinton等^[30]将深度神经网络技术应用于语音识别领域并取得良好效果。音频检索包括基于直方图统计阶段和基于音频指纹DNA的散列检索阶段。特定视频内容的探测分析主要包括视频的特征表示、语义属性分析以及多模融合识别等内容。递归神经网络已经应用于视频的时空特征提取上。语义的属性分析也同样依赖于近期图像物体识别与目标检索方面的进展来进行特定人物、标志、物体等的分析与识别。

3.4 探测结果分析技术

3.4.1 实体资源探测结果分析

对主机、路由器、服务器等实体资源的定位而言,获取可靠的网络拓扑十分重要,网络安全防护、路由器负载均衡、匿名路由等多种因素严重影响了探测结果的可用性,因此,需要对网络探测的结果进行融合分析,以获得高可靠性可用的网络拓扑结果。下面简要介绍网络设备拓扑结构融合分析涉及的主要技术。

1) 网络路径重构技术

网络路径重构技术主要用于解决目标网络拓扑探测结果中由于路由器多个别名导致的路径冗余、目标所在子网的判别等问题。

2) 拓扑分析技术

在将实体资源映射到地理空间时,通常需要获取目标实体资源与周边区域网络节点的连接情况,如时延、跳数等。区域性拓扑分析技术能够从海量的拓扑探测结果中,分析得到目标相关的可用拓扑信息,为后期的实体资源地理位置映射提供数据基础。面向网络可达性与起源变化的拓扑分析技术能够从探测结果中分析得到不同源与不同目标之间的连通性和可达性,可为后期定位算法的设计提供帮助。

3) 拓扑语义标注技术

对网络设备类型、操作系统、性能、重要性等多种属性的准确了解能够为全面掌握网络状态提供帮助,网络拓扑语义标注技术能够根据需求,从不同的层面对网络拓扑进行准确的描述。

3.4.2 虚拟资源探测结果分析

网络空间虚拟资源探测结果分析技术主要围绕文本内容、音视频网站、博客论坛以及其他

互联网平台上的特定内容资源,开展快速发现、关联分析、内容分析与理解等方面的研究,主要技术如下。

1) 特定信息内容快速发现和关联分析技术

特定信息内容快速发现和关联分析技术主要用于快速发现多个网络平台上的特定信息内容,并对其进行重要性、与话题特征的标注或多语种的翻译等。通过对发现的信息内容的关联分析,可以从中挖掘到潜在的信息,得到特定信息内容的传播规律和分布特点等。

2) 特定音频内容的检索与识别技术

特定音频内容的检索与识别技术主要对复杂多变声学环境下的音频进行处理,对海量多语言网络音频数据中的特定语种、说话人和特定内容进行检测,对音频的语种进行识别,以及对互联网海量数据下的音频模板进行匹配,最终实现对包含特定信息的音频检索与识别。

3) 特定视频内容的检测和识别技术

特定视频内容的检测和识别技术通过对视频中的人、物等关键元素的提取,形成针对视频内容的结构化描述,利用有效的视频识别算法,实现对视频中特定场景的检测和识别,并能够根据给定的样例检索出相关视频。

4 网络空间测绘映射层关键技术与研究进展

4.1 映射层技术框架

映射层的主要目的是利用探测层提供的相关数据,将网络实体资源映射到地理空间,将网络虚拟资源映射到社会空间。映射层的技术框架如图3所示。

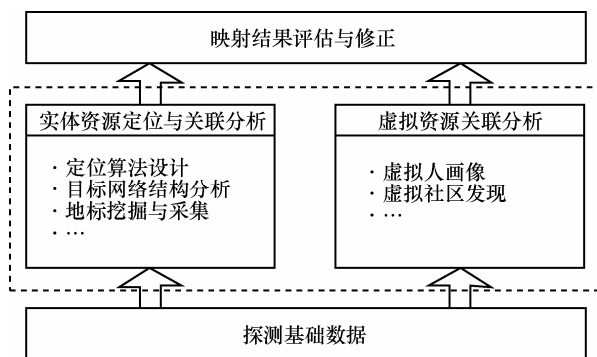


图3 映射层技术框架

其中,实体资源向地理空间映射技术主要包括地标挖掘与采集技术、目标网络结构分析技术、网络实体定位技术等,虚拟资源向社会空间映射技术主要包括虚拟人画像技术和虚拟社区发现技术等。

4.2 实体资源向地理空间映射技术

4.2.1 基本概念和主要步骤

在实体资源向地理空间映射技术中,本文重点讨论具有IP的网络实体资源的地理位置定位技术,涉及的基本概念主要如下。

目标:待定位的IP地址,通常假设其有一个固定的地理位置。

探测源:给目标发送探测分组的主机,一般自身地理位置已知。

地标:自身地理位置已知的IP地址,与探测点不同的是,它不能主动发送与定位有关的数据分组,只能响应探测分组。

定位服务器:用于向探测源发送测量等命令,收集、处理探测源提交的测量数据,执行定位算法,并将最终的定位结果反馈给用户。

实体资源的定位过程如图4所示,可简要概括为:首先,定位服务器在收到待定位目标的IP地址后,根据定位精度需求和定位算法的需要部署探测源,并向各探测源发送相应的测量或查询指令;其次,探测源根据收到的定位服务器的指令,执行相应的操作,如测量与目标或地标的时延或拓扑连接关系等,并将测量的结果提交给定位服务器;最后,定位服务器将探测源提交的信息进行处理后,执行定位算法,将定位的结果存入地标库并输出。

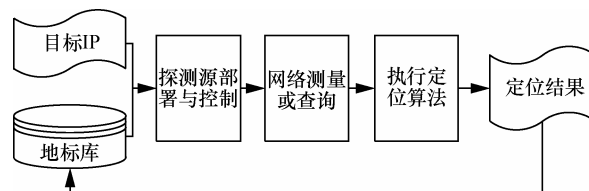


图4 实体资源定位过程

4.2.2 主要相关技术

1) 实体地标获取与评估技术

地标是实现将网络实体映射到地理位置的基准点。在一个区域内数量较大且分布均匀的地标点,既可为该区域内的目标实体定位提供支撑,也可用于验证定位算法的有效性。现有地标获取

方法主要有两类：基于 Web 挖掘的地标获取方法和基于 IP 定位数据库的地标获取方法。

在基于 Web 挖掘的地标获取研究方面，文献[31]通过对特定互联网论坛的挖掘来获得大量的城市级可靠地标，文献[32]提出的方法首先向地图服务发送指定区域的查询请求，地图服务会将该区域内或附近的公司、学术机构及政府机关等组织机构以列表的形式返回，列表中包含这些组织机构详细的地理位置及其主页域名；然后由域名解析出其 IP 地址，从而得到<域名、IP 地址、地理地址>映射关系。该方法可以有效地在 Web 服务发达的地区获取街道级地标，文献[32,33]等都采用此方法采集街道级地标。

在基于 IP 定位数据库的地标获取方面，目前有很多公司在维护和发布将 IP 地址映射到地理位置的数据库，如 MaxMind^[34]、IP2Location^[35]及 Quova^[36]等。此外，WHOIS 数据库^[37]中也有 IP 地址块的地理位置信息，即利用 Whois 查询可得到一个 IP 地址所在地址块在 WHOIS 数据库中记录的位置。基于 IP 定位数据库的地标获取方法，通常选取多个定位数据库中地理位置一致的 IP 用于实验研究。

2) 网络实体定位技术

网络实体定位技术是实现将实体资源映射到地理空间的核心技术。从公开的文献来看，国际上关于网络实体定位技术的研究已经开展了十余年，取得了一些优秀的研究成果；国内的解放军信息工程大学网络空间测绘课题组率先开展了网络空间测绘相关研究，并提出了一些新的网络实体定位方法。现有网络实体定位方法大致可分为 3 类：基于数据库查询的定位、基于数据挖掘的定位和基于网络测量的定位。

基于数据库查询的定位方法由于不需要大量测量，且定位速度快、计算开销小而得到广泛应用。当前互联网上已有多种对外提供查询的 IP 定位数据库，如 Maxmind^[34]、IP2location^[35]、Quoval^[36]、Whois 数据库^[37]等。这些数据库包含 IP 地址及其地理位置的映射关系，并对外提供查询接口。用户只需提交要查询的 IP 地址，通常能够很快获得查询结果。此外，文献[38]通过引入 IP Number 的概念，自主构建了 IPMapping 数据库，

文献[39]利用分布式爬虫等技术提高了 IP2location 数据库的覆盖范围和城市级定位准确率。

基于数据挖掘的定位算法试图从具有组织机构和用户地理位置信息的网站、手机应用等数据来源中，挖掘地理位置与 IP 地址之间的关系。典型的定位方法主要包括 Structon^[33]、Checkin—Geo^[40]等。Structon 算法首先使用网络爬虫从互联网上抓取机构等的主页信息，从中提取电话区号、联系地址等能够表达地理位置的信息，然后通过查询域名将 URL 转化为 IP，从而将 IP 地址与地理位置关联起来，最后将 IP 按照地址段进行迭代推导，扩大可定位的 IP 地址范围，该方法的准确性依赖于所抽取的主页位置信息与 IP 地址的映射关系是否准确，易受服务器托管等情况的影响。Checkin—Geo 算法首先从某种手机端应用获取“用户 ID—地理位置”关系数据，再从对应的 PC 端应用程序中获取“用户 ID—LOG 信息—IP 地址”关系数据，最后利用机器学习等方法掌握用户活动规律，建立起“IP 地址—地理位置”关系以实现目标 IP 定位。该方法可对用户所处的位置进行较为可靠的分析，将用户定位至家庭、工作场所或特定的区域，但需要从手机服务提供商和网络服务提供商处获得关于目标的大量历史数据，因此难以适用于非协作环境。

基于网络测量的定位方法通过测量探测源与目标节点之间的时延（或在此基础上结合拓扑信息），用不同的方法将时延转换为地理距离，以不同的方式对目标节点产生距离约束，然后估计出目标节点的位置。根据产生距离约束的方式不同，可进一步分为基于时延的定位算法、基于概率估计的定位算法和基于拓扑的定位算法 3 类。

基于时延的定位算法主要包括 Shortest Ping^[41]、GeoPing^[41]、CBG (constraint-based geolocation)^[42]、SPRG(segmented polynomial regression approach)^[43]、Geo-RX^[44]、GeoWeight^[45]等，这些方法直接根据时延的特点判断目标距离探测源（地标）的远近，或将时延以某种关系转换为距离约束（通常为目标距离探测源的最远距离），进而估计出目标的可能位置，算法的原理较为简单，但容易受到时延测量不准确的影响，定位精度不高。基于概率估计的典型定位算法主要包括：

LBG(learning-based geolocation)^[46]、ELC (enhanced learning classifier)^[47]、MLE-based Approach(maximum likelihood estimation based approach)^[48]、Spotter^[49]、GBLC^[50]等, 这些算法不假设时延和地理距离符合某种线性关系, 而是通过对大量的时延—距离数据进行统计, 得出时延大小与地理距离的概率统计关系, 能够降低单次时延测量不准确的影响, 但往往定位精度较低, 通常仅能够给出城市级定位。基于拓扑的典型定位算法主要有 SLG(street-level geolocation)^[32]、GeoTrack^[41]、TBG(topology-based geolocation)^[51]、Octant^[52]、GBRF^[53]、GBPF^[54]等, 这些算法使用目标的路径或地标的拓扑信息, 或在时延基础上结合拓扑信息, 然后进行交互验证来实现对目标的定位, 定位精度较高, 但容易受到匿名路由的影响。

现有典型定位算法在测试算法性能时, 通常是在理想的实验网络(如 PlanetLab)或连通性较好的网络环境下进行的, 部分算法的定位误差如表 1 所示。由表 1 可知, 现有定位算法通常只能实现区域级或城市级的定位精度, 仅有少量的算法(如表 1 中的 SLG 算法)能够实现较高精度的定位。

4.3 虚拟资源向社会空间映射技术

4.3.1 技术框架

虚拟资源向社会空间映射技术框架如图 5 所示。

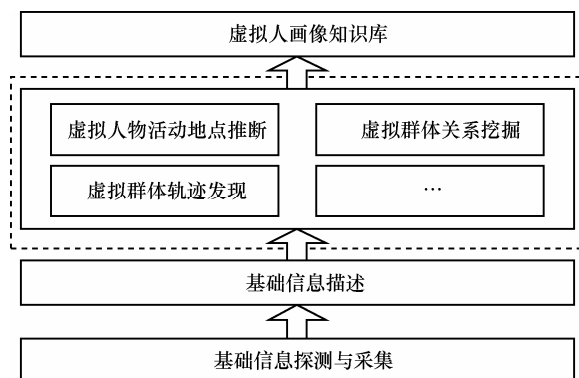


图 5 虚拟资源向社会空间映射技术框架

虚拟资源向社会空间映射技术实现将网络社会成员映射到现实社会成员。基本过程可简要概括为: 首先, 利用 3.3.2 节中的虚拟资源探测技术对网络上与用户相关的多源海量网络日志发起探测或采集, 对得到的来源不同、结构各异的基础信息进行处理, 实现一体化表达; 其次, 利用虚拟人物活动地点推断方法、虚拟群体关系挖掘方法、虚拟群体轨迹发现方法等将网络社会成员与现实社会成员相关联, 实现将网络社会成员向社会成员的映射; 最后, 将映射结果存入虚拟人画像知识库。

4.3.2 主要相关技术

实现虚拟资源向社会空间映射涉及的技术主要包括虚拟人物活动地点推断、虚拟群体关系挖掘和虚拟群体轨迹发现等技术。其中, 虚拟人物

表 1 典型定位算法的误差比较

定位方法	实验数据集	定位误差/km
GeoPing	51 个 PlanetLab 节点, 既作探测源也作目标	中值: 109 最大: 1 714
GeoTrack		中值: 156 最大: 4 335
Octant		中值: 35 最大: 277
CBG		平均: 78, 182 中值: 22, 95
LBG	探测源: 78 个 PlanetLab 节点 地标集: 3 200 个美国境内的路由器 目标集: 12 800 个美国境内的路由器	平均: 112
TBG	探测源: 北美 68 个 PlanetLab 节点 目标: ①非学术 22 个 Sprint 网络中的节点; ②学术 来自美国各大学共 128 个主机	平均: 194, 253
SLG	探测源: 163 个公开可用的 ping 服务器和 136 个 traceroute 服务器 目标集: 88 个 PlanetLab 节点; 居民填写的 72 个<IP, 邮政地址>对; 3 个月的地图查询日志中的<IP, 经纬度>对	中值: 0.69, 2.25, 2.11 最大: 5.24, 8.1, 13.2
GBRF	探测源: 自主部署位于北京市、成都市的 2 个探测源 目标集: 位于河南省的共 1 024 个目标 IP	城市级准确率高于 92%

活动地点推断技术从采集到的数据中综合分析位置相关的信息,以推断虚拟人物的活动地点,尤其在社交网络环境下,人们常常有意或无意地通过交互内容来透露自己的地理信息和短期活动计划,因此,从原始用户产生文本中检索虚拟人物在其中涉及的地理兴趣点(POI)词条,通过设计有效的消歧义算法确定真实的POI,结合POI资源库和用户行为模式,可推断POI的具体地理位置,最终实现对特定人物位置信息的挖掘,如Rae等^[55]提出了一个基于线性条件随机场CRF模型的POI抽取算法,实现了对常规文本中的POI抽取。

虚拟群体关系挖掘技术通过挖掘不同虚拟人物之间的关联关系,实现对虚拟群体的有效描述。

虚拟群体轨迹发现技术通过对群体中虚拟人的活动规律、活动轨迹的关联分析,来获取整个群体的行为特点、活动方式,实现对特定虚拟群体轨迹模式的发现。

5 网络空间测绘绘制层关键技术与研究进展

5.1 绘制层技术框架

网络空间绘制技术在探测结果和映射结果的基础上,将多维的网络空间资源及其关联关系投影到一个低维的可视化空间,构建网络空间的分层次、可变粒度的网络地图,实现对多变量时变型网络资源的可视化。绘制层的技术框架如图6所示。

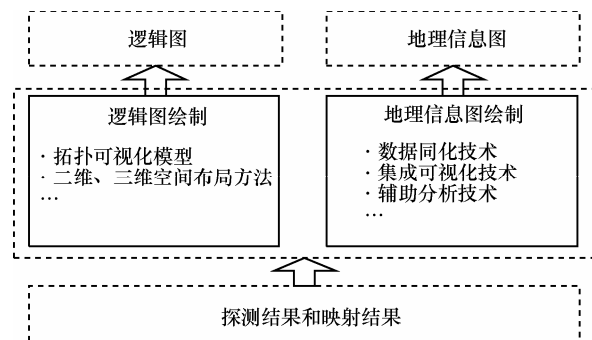


图6 绘制层技术框架

网络空间的绘制主要包括逻辑图绘制和地理信息图绘制,其中逻辑图绘制主要是通过构建拓扑可视化模型,利用二维、三维等空间布局方法

将探测得到的网络拓扑可视化,地理信息图绘制技术利用数据同化技术、集成可视化技术、辅助分析技术等将网络空间资源的网络属性和地理空间属性进行可视化。

5.2 逻辑图绘制技术

网络空间逻辑图的绘制主要指对网络拓扑图的绘制,主要目标是将网络中的节点和连接状况以符合其内在特性的方式完整清晰地展现在用户眼前,从而为人们了解和分析网络空间的整体状况提供直观素材和操作平台。在绘制逻辑图时,往往要求将其性质、度量及模型等体现在可视化结果中,研究重点通常为解决可视区域和逻辑图规模之间的矛盾,以及便于理解的可视化策略的选择或设计。

5.3 地理信息图绘制技术

网络空间地理信息图绘制技术主要实现基于地理空间基础数据的网络空间测绘数据可视化表达,在绘制网络空间地理信息图时,涉及的技术主要包括地理空间和网络空间数据的同化技术、网络空间信息和地理空间信息的集成可视化技术、网络节点辅助分析技术等。其中,地理空间和网络空间数据的同化技术是一种数据处理技术,用于将来自网络空间和地理空间的不同格式、不同性质、不同模型的数据进行融合处理,为可视化提供可用的基础数据;网络空间信息和地理空间信息的集成可视化技术用于集成同化后的网络数据和地理数据,以便于用户理解和后期查询、预测等应用方式进行展示;网络节点辅助分析技术通过对映射结果周边的网络环境和地理环境的分析,来评估映射结果的合理性、可信度和可用性,为修正映射结果提供依据。

6 有待进一步研究的问题

通过上述对网络空间测绘技术研究进展的梳理可知,当前关于网络空间测绘技术的研究已经取得了一定进展,但是,实现对网络空间资源的精确测绘,仍有一些问题亟待解决,具体如下。

1) 如何实现对内网和暗网资源的探测?

分布或存储在内网和暗网中的资源丰富,是网络空间资源态势的重要组成部分。但是,内网中的资源或特定目标往往处于防火墙、VPN等多

种技术保护之下,暗网中的资源无法通过标准的搜索引擎进行查找,如何在不影响设备正常工作、不侵犯用户个人隐私的情况下,全面掌握内网资源或快速发现内网和暗网中的特定目标是网络空间探测技术亟待解决的问题之一。

2) 如何实现对实体资源的高精度地理位置映射?

服务器、路由器、网络主机等网络实体资源是网络空间资源的重要组成部分,是网络空间测绘技术关注的重要对象之一。但上述网络实体资源自身通常不具备定位功能,且无法向外界通告自身位置,其网络唯一标识IP地址也不包含位置信息。现有基于IP的网络实体定位方法虽然取得了很大进展,但大多仅能给出城市级的位置信息,且可靠性难以保证,因此,定位精度和定位结果的可靠性仍有待提高。

3) 如何利用大数据分析技术实现对网络虚拟人物/社团的精确画像?

大数据时代已然来临,网络上与虚拟人物相关的信息规模庞大,但这些数据往往类型多样,既包含结构化的数据,又包含非结构化数据,且数据的真伪难辨。如何从海量数据中挖掘、分析与虚拟人物/社团相关的信息,进而将网络虚拟人物与其社会身份准确关联起来,是网络空间映射技术需要解决的问题之一。

4) 如何将海量数据进行可视化?

网络空间资源类型多样,规模庞大,地理空间数据同样多源异构,如何将2种空间数据进行格式转换、时空基准统一及信息融合,实现对网络空间数据和地理空间数据的一体化实时动态可视化表达,并能够基于地理空间基础数据对网络节点的地理位置映射精度和可靠性进行分析,是网络空间绘制技术需要解决的问题之一。

7 结束语

本文对网络空间测绘技术研究进展进行了综述。首先,给出了网络空间测绘技术的定义和主要研究内容,提出了网络空间测绘技术框架,主要包括探测层、映射层和绘制层3个层次;其次,分别介绍了3个层次涉及的相关技术及其研究进展,其中,探测层主要包括探测通道与平台技术、

多种探测技术和探测结果分析技术,映射层主要包括实体资源向地理空间映射技术和虚拟资源向社会空间映射技术,绘制层主要包括逻辑图绘制技术和地理信息图绘制技术;最后,给出了网络空间测绘技术目前亟待解决的几个关键问题。网络空间测绘作为网络空间安全和应用的重大共性关键基础技术,必将越来越多地受到学术界和产业界的关注。

参考文献:

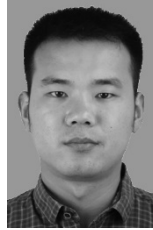
- [1] [EB/OL]. https://en.wikipedia.org/wiki/Plan_X.
- [2] [EB/OL]. <http://mobile.businessinsider.com/nsa-treasure-map-project-2014-9>.
- [3] PAPAPETROU O, SAMARAS G. Ipmicra: toward a distributed and adaptable location aware Web crawler[C]//The 8th East European Conference on Advances in Databases and Information Systems. 2004.
- [4] CAMBAZOGLU B B, TURK A, KARACA E, et al. SE4SEE: a grid-enabled search engine for south-east Europe[C]//The International Conference on Hypermedia and Grid Systems. 2005: 223-227.
- [5] SINGH A, SRIVATSA M, LIU L, et al. Apoidea: a decentralized peer-to-peer architecture for crawling the world wide Web[C]//International Workshop on Distributed Information Retrieval. 2003: 126-142.
- [6] SPRING N, MAHAJAN R, WETHERALL D. Measuring ISP topologies with rocketfuel[J]. ACM Sigcomm Computer Communication Review, 2002, 32(4): 133-145.
- [7] KATZ-BASSETT E, MADHYASTHA H, ADHIKARI V, et al. Reverse traceroute[C]//The 7th Usenix Symposium on Networked Systems Design and Implementations (NSDI). 2010: 219-234.
- [8] Nmap: a free network mapping and security scanning tool[EB/OL]. <http://nmap.org/>.
- [9] DURUMERIC Z, WUSTROW E, HALDERMAN J A. ZMap: fast internet wide scanning and its security applications[C]//The 22nd Usenix Security Symposium. 2013: 605-620.
- [10] LI J. Network topology discovery based on SNMP[C]//The 9th International Conference on Computational Intelligence and Security(CIS). 2013: 194-199.
- [11] YIN J B, LI Y M, WANG Q, et al. SNMP-based network topology discovery algorithm and implementation[C]//The International Conference on Fuzzy Systems and Knowledge Discovery(FSKD). 2012: 2241-2244.
- [12] MARCHETTA P, MÉRINDOL P, DONNET B, et al. Topology discovery at the router level: a new hybrid tool targeting ISP networks[J]. Selected Areas in Communications, 2011, 29(9): 1776-1787.
- [13] MARCHETTA P, MÉRINDOL P, DONNET B, et al. Quantifying and mitigating IGMP filtering in topology discovery[C]//The Conference on Global Communications (GLOBECOM). 2012: 1871-

- 1876.
- [14] MARCHETTA P, PESCAPE A. DRAGO: detecting, quantifying and locating hidden routers in traceroute IP paths[C]//The 32nd IEEE International Conference on Computer Communications (INFOCOM). 2013: 3237-3242.
- [15] GUNES M H, SARAC K. Resolving IP aliases in building traceroute-based Internet maps[J]. *IEEE/ACM Transactions on Networking (ToN)*, 2009, 17(6): 1738-1751.
- [16] SHERRY J, KATZ-BASSETT E, PIMENOVA M, et al. Resolving IP aliases with prespecified timestamps [C]//The 10th ACM SIGCOMM Conference on Internet Measurement. 2010: 172-178.
- [17] MARCHETTA P, PERSICO V, PESCAPÈ A. Pythia: yet another active probing technique for alias resolution [C]//The 9th ACM Conference on Emerging Networking Experiments and Technologies. 2013: 229-234.
- [18] SPRING N, MAHAJAN R, WETHERALL D, et al. Measuring ISP topologies with rocketfuel[J]. *IEEE/ACM Transactions on Networking*, 2004, 12(1): 2-16.
- [19] MADHYASTHA H V, ISDAL T, PIATEK M, et al. iPlane: an information plane for distributed services[C]//The 7th Symposium on Operating Systems Design and Implementation. 2006: 367-380.
- [20] FELDMAN D, SHAVITT Y. Automatic large scale generation of internet pop level maps[C]//The Conference on Global Communications (GLOBECOM). 2008: 1-6.
- [21] SHAVITT Y, ZILBERMAN N. Geographical internet pop level maps[C]//The International Conference on Traffic Monitoring & Analysis. 2012: 121-124.
- [22] SHAVITT Y, ZILBERMAN N. Improving IP geolocation by crawling the Internet PoP level graph[C]//The Conference on IFIP Networking. 2013: 1-9.
- [23] KOHNO T, BROIDO A, CLAFFY K C. Remote physical device fingerprinting[J]. *IEEE Transactions on Dependable and Secure Computing*, 2005, 2(2): 93-108.
- [24] FINK R. A statistical approach to remote physical device fingerprinting[C]//The Conference on Military Communications. 2007: 1-7.
- [25] DANISMAN T, ALPKOÇAK A. Feeler: emotion classification of text using vector space model[C]//The AISB Convention on Communication, Interaction and Social Intelligence. 2008: 53-59.
- [26] TANG B, WU Y, JIANG M, et al. Recognizing and encoding disorder concepts in clinical text using machine learning and vector space model[C]//Working Notes for CLEF 2013 Conference. 2013: 23-26.
- [27] ALSALLAL M, IQBAL R, AMIN S, et al. Intrinsic plagiarism detection using latent semantic indexing and stylometry[C]//The 6th International Conference on Developments in eSystems Engineering (DeSE). 2013: 145-150.
- [28] YE H J F, TAN Y S, LEE C H. Topic detection and tracking for conversational content by using conceptual dynamic latent dirichlet allocation[J]. *Neurocomputing*, 2016.
- [29] CHEN Y, AMIRI H, LI Z, et al. Emerging topic detection for organizations from microblogs[C]//The 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2013: 43-52.
- [30] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.
- [31] ZHU G, LUO X, LIU F, et al. An algorithm of city-level landmark mining based on Internet forum[C]//The 18th International Conference on Network-Based Information Systems(NBIS). 2015: 294-301.
- [32] WANG Y, BURGNER D, FLORES M, et al. Towards street-level client-independent IP geolocation[C]//The 8th Conference on Networked Systems Design and Implementation (NSDI). 2011: 27-27.
- [33] GUO C, LIU Y, SHEN W, et al. Mining the web and the Internet for accurate IP address geolocations[C]//The 28th IEEE Conference on Computer Communications (INFOCOM). 2009: 2841-2845.
- [34] [EB/OL]. <https://www.maxmind.com/zh/home>.
- [35] [EB/OL]. <http://www.ip2location.com>.
- [36] [EB/OL]. <https://www.neustar.biz/services/ip-intelligence>.
- [37] [EB/OL]. <https://www.whois.net>.
- [38] 陈鹏. 网络实体地理定位 IP Mapping 系统设计与实现[D]. 长沙: 国防科学技术大学, 2008.
- CHEN P. Design and implementation of network entity geolocation IP Mapping system[D]. Changsha: National University of Defense Technology, 2008 .
- [39] 李威. 基于 IP 地址的网络实体地理位置定位技术研究及实现[D]. 北京: 北京交通大学, 2008.
- LI W. Research and implementation of the network entities location positioning technology based on the IP address[D]. Beijing: Beijing Jiaotong University, 2008.
- [40] LIU H, ZHANG Y, ZHOU Y, et al. Mining checkins from location-sharing services for client-independent IP geolocation[C]//The 33th IEEE Conference on Computer Communications (INFOCOM). 2014: 619-627.
- [41] 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.
- [42] 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.
- [43] DONG Z, PERERA R D W, CHANDRAMOULI R, et al. Network measurement based modeling and optimization for IP geolocation[J]. *Computer Networks*, 2012, 56(1): 85-98.
- [44] LAKI S, MÁTRAY P, HÁGA P, et al. A model based approach for improving router geolocation[J]. *Computer Networks*, 2010, 54(9): 1490-1501.
- [45] ARIF M J, KARUNASEKERA S, KULKARNI S. GeoWeight: Internet host geolocation based on a probability model for latency measurements[C]//The 33rd Australasian Conference on Computer Science. 2010: 89-98.
- [46] ERIKSSON B, BARFORD P, SOMMERSY J, et al. A learning-based approach for IP geolocation[C]//The 11th International

- Conference on Passive and Active Measurement. 2010: 171-180.
- [47] MAZIKU, HELLEN, SHETTY, et al. Enhancing the classification accuracy of IP geolocation[C]//The Conference on Military Communications. 2012: 1-6.
- [48] ARIF M J, KARUNASEKERA S, KULKARNI S, et al. Internet host geolocation using maximum likelihood estimation technique[C]//The 24th IEEE International Conference on Advanced Information Networking and Applications. 2010: 422-429.
- [49] LAKI S, MÁTRAY P, HÁGA P, et al. Spotter: a model based active geolocation service[C]//The 30th IEEE Conference on Computer Communications (INFOCOM). 2011: 3173-3181.
- [50] ZHU G, LUO X, LIU F, et al. City-level geolocation algorithm of network entities based on landmark clustering[C]//The 18th International Conference on Advanced Communication Technology (ICACT). 2016: 306-309.
- [51] KATZ-BASSETT E, JOHN J P, KRISHNAMURTHY A, et al. Towards IP geolocation using delay and topology measurements[C]//The 6th ACM Sigcomm Conference on Internet Measurement. 2006: 71-84.
- [52] WONG B, STOYANOV I, SIRER E G. Octant: a comprehensive framework for the geolocalization of Internet hosts[C]//The 4th Usenix Symposium on Networked Systems Design & Implementation. 2007: 313-326.
- [53] ZHAO F, SONG Y, LIU F, et al. City-level geolocationbased on routing feature[C]//The 29th International Conference on Advanced Information Networking and Applications. 2015: 414-419.
- [54] CHEN J, LIU F, WANG T, et al. Towards region-level IP geolocation based on the path feature[C]//The 17th International Conference on Advanced Communication Technology (ICACT). 2015: 468-471.
- [55] RAE A, MURDOCK V, POPESCU A, et al. Mining the web for

points of interest[C]//The 35th International ACM Sigir Conference on Research and Development in Information Retrieval. 2012: 711-720.

作者简介:



赵帆 (1989-), 男, 江苏徐州人, 解放军信息工程大学博士生, 主要研究方向为网络空间测绘与实体定位。



罗向阳 (1978-), 男, 湖北荆门人, 解放军信息工程大学副教授、博士生导师, 主要研究方向为网络空间测绘与实体定位、多媒体安全。



刘粉林 (1964-), 男, 江苏溧阳人, 解放军信息工程大学教授、博士生导师, 主要研究方向为网络与信息安全。