软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2014,25(7):1527−1540 [doi: 10.13328/j.cnki.jos.004621]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

IP 定位技术的研究*

王占丰[1,3], 冯 径[1], 邢长友[2], 张国敏[2], 许 博[2]

[1](解放军理工大学 气象海洋学院,江苏 南京 211101)

[2](解放军理工大学 指挥信息系统学院,江苏 南京 210007)

[3](93615 部队,天津 300301)

通讯作者: 王占丰, E-mail: hehengw@hotmail.com

摘 要: IP 定位技术就是确定 Internet 中 IP 设备的地理位置,它可以帮助网络应用改善性能、提高安全性以及提供新的服务.首先概述了 IP 定位技术的基本概念和应用情况;然后,将现有定位算法分为独立于客户端和基于客户端两类定位算法,并对每一类算法中的典型算法进行了具体分析,讨论了隐私保护技术和新技术的影响;最后,对现有的 IP 定位算法进行了综合对比,指出了 IP 定位技术的研究方向.

关键词: 网络测量;IP 定位;网络性能;基准节点

中图法分类号: TP393

中文引用格式: 王占丰,冯径,邢长友,张国敏,许博.IP 定位技术的研究.软件学报,2014,25(7):1527−1540. http://www.jos.org.cn/1000-9825/4621.htm

英文引用格式: Wang ZF, Feng J, Xing CY, Zhang GM, Xu B. Research on the IP geolocation technology. Ruan Jian Xue Bao/Journal of Software, 2014,25(7):1527−1540 (in Chinese). http://www.jos.org.cn/1000-9825/4621.htm

Research on the IP Geolocation Technology

WANG Zhan-Feng[1,3], FENG Jing[1], XING Chang-You[2], ZHANG Guo-Min[2], XU Bo[2]

[1](Institute of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101, China)

[2](College of Command Information Systems, PLA University of Science and Technology, Nanjing 210007, China)

[3](93615 PLA Troops, Tianjin 300301, China)

Corresponding author: WANG Zhan-Feng, E-mail: hehengw@hotmail.com

Abstract: IP geolocation aims at determining the geographic location of an Internet host, which can improve the performance and security of the Internet application, and bring about novel services. This paper firstly illustrates the concept and applications of the IP geolocation, and then categorizes current typical IP address geolocation algorithms into two classes, namly, client-independent geolocation algorithms and client-dependent geolocation algortithms. Next, the main ideas of the representative algorithms of each class are illustrated, and the privacy protection techniques and the influence of new techniques are discussed. Finally, a comprehensive comparison is made on the IP geolocation algorithms and systems, and the future trends of the IP geolocation are discussed.

Key words: network measurement; IP geolocation; network performance; landmark

IP 定位技术,简而言之,是通过设备的 IP 地址来确定其地理位置.近年来,基于地理位置的网络应用层出不穷,主要包括定向广告(targeted advertisement)、社交网络、网络安全、性能优化等[1−4].许多研究机构和学者围绕 IP 定位技术从定位精度、隐私保护以及应用等不同角度展开了研究.著名的 Internet 测量组织 CAIDA 在美国国土安全部的资助下,对各种 IP 定位算法进行了分析[5].IETF 成立了专门的工作组来讨论 Internet 定位技术的标准、隐私保护等相关问题,并提出了相应的草案[6].国际互联网组织 W3C 更是在下一代网络标准中将定位

* 基金项目: 国家重点基础研究发展计划(973)(2012CB315806); 国家自然科学基金(61371119, 61379149, 61103225, 61070173)
收稿时间: 2013-09-21; 修改时间: 2014-02-17; 定稿时间: 2014-04-08

© 中国科学院软件研究所 http://www.jos.org.cn

服务作为一个标准接口写入了 HTML5 规范[7].

IP 定位的基本原理是,利用 IP 设备的名字、注册信息或时延信息等来估计其地理位置.定位算法设计的基本原则是:在保证定位精度的前提下,尽量减少测量开销,同时兼具良好的扩展性,并能保护用户隐私.最初的定位算法通过向 DNS 服务器查询或者挖掘隐含在主机名中的信息来推测 IP 设备的地理位置[8−10].之后,一些定位算法根据时延与地理距离之间的线性关系来估测主机位置,并通过拓扑信息来减小定位误差[11−13].近年来,基于概率的定位算法重新成为一个研究热点,通过寻找时延与地理距离的分布规律来进行定位[14−17].虽经不断改进,但这两类算法都不能精确地定位,因此,一些综合的定位算法使用了上述两类方法来进行交叉验证以提高精度[1,13].

IP 定位算法可以按照是否需要客户端的支持、定位原理等不同标准进行分类,本文将其分为独立于客户端的定位算法和基于客户端的定位算法.这两类算法各有优劣:独立于客户端的定位算法主要借助推测、网络测量等方法推断目标主机位置;基于客户端的定位算法精度较高,但是往往要借助 GPS、蜂窝基站、WiFi 接入点等基础设施.如今,伴随着社交网络的流行,用户地理位置被公布出来,一方面促进了好友间的交流,另一方面也带来用户对于隐私泄露的担忧.此外,IPv6 网络的大面积部署和位置标识/身份标识分离协议(locator/ID separation protocol,简称 LISP)等新型协议的提出,也为 IP 定位技术的发展带来了新的机遇和挑战.

本文第 1 节对 IP 定位技术进行概述.第 2 节、第 3 节对典型的独立于客户端的定位算法和基于客户端的定位算法进行详细分析.第 4 节分析隐私保护技术以及网络新技术对定位的影响.第 5 节综合分析各种 IP 定位算法和面临的挑战.第 6 节总结全文.

1 IP 定位概述

1.1 基本概念

简单地说,IP 定位技术就是为确定 IP 设备地理位置所采用的技术.在 IP 定位系统或算法中,一般包括 4 个要素:定位服务器、测量节点、待定位节点和基础设施,如图 1 所示.

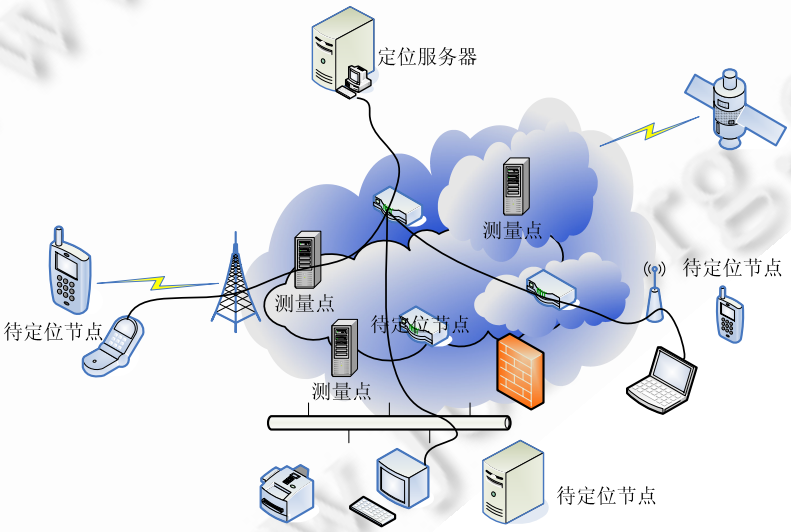


Fig.1 The scene of IP geolocation

图 1 IP 定位场景

- 定位服务器:通过协调测量节点对待定位节点进行定位,并收集、处理定位信息向最终用户提供定位服务;
- 测量点:在定位服务器的控制下,通过网络测量或信息查询来获得待定位节点的时延、路由、位置信息

© 中国科学院软件研究所 http://www.jos.org.cn

等,并提交给定位服务器;

- 待定位节点是具有 IP 地址的设备,包括计算机、手机、路由器等;
- 基础设施:泛指用于定位的协议和设备,如网络协议、服务接入点、GPS 定位系统、蜂窝基站等.

IP 定位的基本过程就是通过设备的 IP 地址测量获得其属性信息,在分析属性信息的基础上获得 IP 设备的地理位置.这个属性信息既可以是主机的名字、IP 地址本身,也可以是 IP 设备与已知定位位置 IP 设备的时延以及连接关系等.

按照是否需要客户端的支持,IP 定位技术可以分为独立于客户端的定位算法和基于客户端的定位算法.按照定位原理,可以分为基于推测的定位算法、基于时延的定位算法以及同时使用这两类技术综合定位算法这 3 类.基于推测的定位方法通过网络注册信息、主机名等信息来推测 IP 地址的地理信息;基于测量的定位系统或方法通过时延与地理之间线性关系来估测主机的位置,并通过拓扑信息来提高预测精度;综合的定位方法则使用上述两种方法来综合提高预测精度.为了描述方便,本文主要按照是否需要客户端的支持对现有定位算法进行分类,对于算法较多的非基于客户端的定位算法,进一步按照定位原理再分类讨论.

1.2 IP定位技术的应用

目前,IP 定位技术的应用主要分为 4 类:网络性能优化、网络安全、社交网络和定向广告.

基于 IP 定位的网络性能优化主要是通过定位服务为用户选择通信距离最短的链路,从而达到减小网络开销、提升网络性能的目的.如在 P2P 网络中,通过为对等节点选择地理位置近的节点通信,以减少跨域流量提高下载速度[18].Cheng 等人利用地理位置信息来控制地理距离短的移动设备进行通信,从而减少数据传输中能量的损失[19].

基于 IP 定位的网络完全应用主要是通过用户的位置信息来对其身份进行验证,可以应用于入侵检测和用户访问控制.在入侵检测中,通过将用户的 IP 地址和地理位置相关联,在一定程度上减少 IP 哄骗攻击.在用户接入控制中,则可以根据分析用户身份的真实性并依据访问控制策略,响应或拒绝用户的通信请求[20].

基于 IP 定位的社交网络主要是通过将社交网络成员的位置或者运动轨迹与他们的活动、兴趣进行关联,向用户进行好友推荐或提供特色服务.如,Facebook 就可以为用户上传的图片自动提供位置信息[21];而 QQ、飞信、微信等通信工具则会依据地理位置推荐好友.

定向广告是指网络服务商利用网络追踪技术(如 Cookies)搜集整理用户信息,按年龄、性别、职业、爱好、收入、地域分类储存用户的 IP 地址,然后向不同类型的用户发送内容不同的广告.由于这类广告针对性强、更加有效,许多浏览器和内容提供商都提供这项服务,如 FireFox、Facebook、亚马逊、腾讯等.

在上述 4 类应用中,前两类属于经典应用,已得到了较好的研究和应用.伴随着社交网络的不断发展和网络服务越来越重视用户的体验,后两者将得到继续发展.在这一过程中,网络服务提供者需要得到更加准确的用户位置信息,而用户在希望得到好的服务的同时尽可能地隐藏自己的真实身份和位置信息.这一矛盾将推动 IP 定位技术围绕定位准确性和隐私保护两个方面进行发展.

1.3 IP定位技术与其他定位技术的关系

定位技术是一个重要的研究领域,它广泛地应用于军事、通信、科研、交通等各个领域.目前,常见的定位技术主要有 GPS 卫星定位、蓝牙定位、WiFi 网络定位、GPRS/CDMA 移动通信技术定位、雷达定位技术、无线传感器定位等.这些定位技术的基本原理是:探测设备直接向待定位的目标发射电磁波,在客户端的协作或非协作下,通过测量与目标的时延来解算目标的位置信息.

相对于上述定位技术,IP 定位技术的应用场景更为复杂.在 Internet 中,迂回路径和时延抖动普遍存在,使得测量点无法采用前面的定位方法直接进行定位,而必须考虑 Internet 的特点.IP 定位技术不仅具有其独特性,同时它又是一种综合的定位技术,前述的定位技术可以作为 IP 定位技术的一部分或者辅助手段.

© 中国科学院软件研究所 http://www.jos.org.cn

2 独立于客户端的定位算法

独立于客户端的定位算法根据其定位原理可以分为基于推测的定位算法、基于时延的定位算法以及综合定位算法这 3 类.

2.1 基于推测的定位算法

基于推测信息的定位算法一般通过查询 Whois 数据库等来获取该 IP 地址的主机名,所在街道或者通过 IP 地址段地理位置来推测该 IP 设备的位置,典型的算法有 IP2LL[9],NetGe[10],GeoTrack[11],VisualRoute[22],Neotrace[23],GTrace[24],Quova[25],MaxMind[26],EdgeScape[27],TraceWare[28],Software77[29],IPligence[30],HostIP[31],IPInfoDB[32]等.这类算法大致可以分为 3 类:第 1 类是直接查询 Whois 数据库来推测主机位置信息;第 2 类是通过测量主机名并结合数据库信息来推断主机地理位置;第 3 类是通过网络结构和数据库信息来推断主机地理位置.

在上述 3 类定位算法中,第 1 类算法提出的最早也最为简单,代表算法包括 IP2LL[9]和 NetGeo[10]等.这类算法准确性差,主要是由以下两个方面造成的:首先,Whois 数据库中的地址信息非常不准确或已经过时,同时,多个服务器对同一个 IP 地址块的地理位置记录可能不一致;其次,一个大的地址块可能会被分配给一个具有广泛地理位置分布的组织,而服务器只会为整个 IP 地址块记录一个地理位置.

基于主机名来推断地理位置的定位算法包括 GeoTrack[11],VisualRoute[22],GTrace[24],Neotrace[23]等,以 GeoTrack 为例,Venkata 等人认为,在主机名字中可能包含了不同粒度的地理位置信息.如,corerouter1.SanFrancisco.cw.net 表明该主机位于 San Francisco,www.state.ca.us 表明主机位于美国的加利福尼亚州.因此,通过挖掘这些信息可以推断主机的位置.该算法共分为 3 步:首先,采用 Traceroute 工具确定测量主机与被测主机之间的网络路径,尽可能地测量出中间节点的 DNS 名字;然后,按照路径依次提取出中间节点的位置信息,从而推断到达被测节点的地理路径;最后,GeoTrack 将路径上最靠近被测主机的可识别路由器的位置作为主机的位置.显然,该算法的精度主要取决于路由器 DNS 名字的定位精度以及目标主机与最后一跳路由器的距离.

GeoCluster[11]是通过网络结构和数据库信息来推断主机地理位置的典型定位算法.在 Internet 中,地址划分和路由是按照层次结构组织的,因此,路由器中的 IP 前缀(address prefix,简称 AP)信息可以作为拓扑簇划分依据.然而,一个大的 ISP 或组织可能覆盖很大一个区域,为了进行准确定位,需要进一步将大的 IP 簇划分为粒度更加适合的簇.为此,GeoCluster 引入了一个阈值,如果位于同一个簇内的 IP 地址数目大于阈值,则成为一个簇;否则,将这个 IP 地址块平均划分成两块,依此不断迭代,从而将 IP 地址块划分为不可再分的 IP 地址簇集合.假设有一个 IP 地址簇为 128.127.126.0/24 网段,根据映射信息得知:有 10 个 IP 地址位于西雅图,而另一个 IP 地址位于波士顿,故可以推断整个簇的地理位置位于西雅图.在 GeoCluster 算法中,最为重要的是 IP 地址到地理位置的映射,而这些信息可以从 Hotmail,bCentral,FooTV 等网站的注册信息中获得.

Quova,MaxMind,EdgeScape,TraceWare,Software77,Ipligence,HostIP,IPInfoDB 等基于数据库的商业定位系统出于商业利益的需要,没有公开其技术细节.这些系统综合各种方法来收集、获得用户信息,甚至通过购买商业网站中的注册信息、ISP 网络拓扑结构和路由策略来获得用户的位置信息,因而具有相对较高的定位精度.

总体来说,基于推测的定位方法不需要大量测量,具有计算开销较小、定位速度快的特点,但是对于一个大型的 ISP 或者地域分散的组织,这种方法的定位精度不高.因此,基于推测的定位算法适用于定位精度要求不高的网络应用.

2.2 基于时延的定位算法

基于时延的定位方法通过测量目标主机到测量点的时延来估测主机位置,为了提高定位精度,往往结合网络拓扑信息来进行定位,典型的定位算法和系统有 Geoping[11],Shortest Ping[3],Constraint-based Geolocation (CBG)[12],Topology-based Geolocation (TBG)[3],NBIGA[14],Posit[33],Geo-RX[34],Spotter[17]等.与基于推测的定位算法相比,这类定位算法具有较为完善的数学基础,可分为基于空间理论的定位算法和基于概率估计的定位算法两类.

© 中国科学院软件研究所 http://www.jos.org.cn

2.2.1 基于空间理论的定位算法

基于空间理论的定位算法利用时延与地理距离的线性映射关系来估计地理距离的远近.按照算法提出的时间顺序,依次分析 Geoping,Shortest Ping,CBG,TBG,Geo-RX[34]这 5 种算法的基本思想和原理.

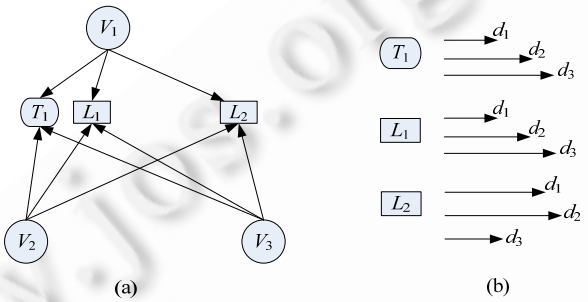
Geoping 算法的基本假设是:到同一组主机时延相近的主机,其地理位置也相近.该算法将所有网络节点划分为目标主机 $T=\{T_1,T_2,\ldots,T_n\}$、测量点 $V=\{V_1,V_2,\ldots,V_n\}$和基准节点 $L=\{L_1,L_2,\ldots,L_n\}$这 3 类.

为确定目标主机的位置,所有的测量点首先向基准节点发起测量,获得由时延构成的一组时延向量 $D_{V_i}=(d_1,d_2,\ldots,d_n)$作为指纹信息.在确定某台主机 T_i 的地理位置时,所有的测量点向目标主机发起主动测量,获得其时延向量 D_{T_i} ,然后比较 D_{T_i} 与 D_{V_i} 的相似性,选择时延向量与 D_{T_i} 最接近节点的地理坐标作为主机 T_i 的地理位置.

在匹配中,向量之间的相似度是通过 N 维时延空间中的欧氏距离来比较的:

$$\hat{D}=\| D_{T_i}-D_{V_i} \|=\left(\sum_{k=1}^{M}(d_{ik}-d_{jk})^2\right)^{1/2},\forall i,j=1,2,\ldots,N \tag{1}$$

图 2 中给出了 Geoping 算法的一个实例,包括 3 个测量节点、2 个基准节点和一个目标主机.通过比较可以看出:主机 T_1 与 L_1 十分接近,因此可以采用 L_1 的坐标近似作为主机 T_1 的坐标.



测量点(vantage points,简称 V)、目标主机(target host,简称 T)与基准节点(landmarks,简称 L)

Fig.2 The main idea of Geoping algorithm

图 2 Geoping 算法基本思想

Shortest Ping 算法的思想是:通过测量基准节点到被测节点的时延,用到被测主机时延最小的基准节点位置作为被测主机的地理位置.可以视为对 Geoping 算法的一种近似和简化,然而其预测精度和许多比它复杂的算法精度相差不大[3].Geoping 与 Shortest Ping 两种算法的精度都受到基准节点数目及其位置的影响,基准节点距离被测节点越近,则其预测误差越小.Hu 等人[35]发现:只有当基准节点与被测节点距离较近时,才能获得较高的定位精度,因此提出了选择邻近节点对 Shortest Ping 和 CBG 算法进行改进.

与 Geoping 不同的是,CBG 不用基准节点的地理位置来近似主机的位置,而是采用一种类似于三角定位的方法来确定被测主机的位置.CBG 假设时延与地理距离之间存在一种线性关系,如公式(2)所示:

$$d_i=m_i\cdot x+b_i \tag{2}$$

其中,d_i 代表基准节点 L_i 到被测节点的地理距离,x 代表两种节点之间的时延,b_i 代表基准节点的某种本地时延(如拥塞时延等).这样就可以概率地估计被测主机到基准节点的距离,当基准节点的数目大于 3 时,则可以形成一个交叠区域,从而确定被测主机的位置,如图 3 所示.

从定位原理上看:Geoping 是用基准节点位置信息来近似代替待定位节点的位置;而 CBG 算法则是由基准节点再发起测量,将待定位节点缩小到一个更小的范围内,其精度要高于 Geoping 算法.然而测量结果表明:在 Internet 中,时延与地理距离的线性关系并不明显.考虑到时延抖动的影响,要提高预测精度,需要进一步改进算法.

Ethan 等人[3]发现:基于基准节点的定位算法,其定位精度并不比简单的定位算法精度高出很多;而基准节

© 中国科学院软件研究所 http://www.jos.org.cn

点与待定位节点的远近,则对于定位精度影响较大.结合传感器网络定位的原理,他们提出了基于拓扑的定位算法(topology-based geolocation,简称 TBG).该算法将测量的网络路径信息作为约束目标主机和中间节点位置的约束条件,然后确定待定位节点和中间节点所在的位置区间.

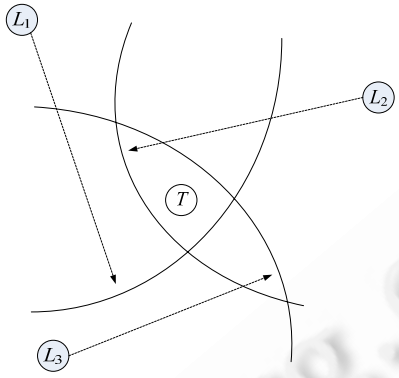


Fig.3 The principal of CBG algorithm

图 3 CBG 算法定位原理

设待定位节点集为 $X=[x_1,x_2,\ldots,x_m]$,基准节点集为 $L=[l_1,l_2,\ldots,l_n]$,则所有待定位节点与基准节点的时延必须符合以下两个约束:

(1) 时延强约束

$$d(l_i,x_j)\leqslant c_{ij} \tag{3}$$

即,如果已知待定位节点 x_j 到基准节点 l_i 的往返时延,则两者 $d(l_i,x_j)$ 必须小于光在光纤中经过往返时延所传输的距离 c_{ij},称其为时延强约束,将所有强约束条件表示为 C_d.

(2) 链路时延弱约束

$$d(x_i,x_j)=h_{ij}+e_{ij} \tag{4}$$

在TBG算法中,通过测量相邻两个节点间的时延来估测它们的距离 h_{ij},但是由于各种因素的影响,h_{ij} 并不准确,将其与真实距离之间的误差表示为 e_{ij},则称为链路时延弱约束,所有的弱约束条件表示为 C_l.

显然,e_{ij} 越小,则表示定位精度越高.TBG 定位算法将定位问题转化为一个带约束的非凸优化问题.TBG 算法的优点是:① 充分利用了距离基准节点越近则定位越准确的优点;② 利用中间节点地理位置来检验到目标节点路径的正确性;③ 允许通过迭代达到全局的一致性.以 Abilene 和 Sprint 网络为例,TBG 可以消除平均误差分别为 40%和 70%,其平均预测精度为 22miles.然而,与其他基于时延的定位算法一样,该算法无法避免时延与地理距离非线性关系对定位精度的影响.

与上述三者不同的是,Geo-RX[9]算法通过建立准确的时延模型来提高定位精度.它将时延分为传播时延 D_{pg}、传输时延 D_{tr}、排队时延 D_q、处理时延 D_{pc} 以及应用层响应测量的时延 D_g.同时,该模型还考虑到测量中正向路径 H_{fw} 与逆向路径 H_{bw} 的差异,如公式(5)所示:

$$d(i,j)=D_g+\sum_{k}^{H_{fw}+H_{bw}}(D_{pc}^k+D_q^k+D_{tr}^k+D_{pg}^k(n_{k-1},n_k)) \tag{5}$$

最终,通过传播时延来确定主机的位置.为了提高预测精度,对全局还进行了优化.

这种算法虽然精度高于 Geoping,但是计算复杂度较高.

2.2.2 基于概率估计的定位算法

基于概率估计的定位算法通过统计时延大小与地理距离的分布关系来进行定位.这种定位方法在网络状况稳定的情况下,可以不受时延与地理距离非线性关系的影响,典型的算法包括 NBIGA[14],Posit[31],Spotter[17]等.

Eriksso 等人[14]认为,CBG 等算法的误差主要是由不同地区的测量密度差异而引起的.因此,他们提出了一

© 中国科学院软件研究所 http://www.jos.org.cn

种基于贝叶斯估计的定位算法(naive Bayes IP geolocation algorithm,简称 NBIGA),通过使用更多的测量数据来降低由于测量数据不足以及不规则网络路径引起的误差,从而将 IP 定位转化为机器学习的分类问题.其基本思想是,将每个待定位的IP主机依据测量结果划分为不同的地理位置集合.NBIGA算法从基准节点向待定位节点发起测量,获得跳数和时延 $M=[m_1,m_2,\ldots,m_n]$,而将所有节点划分的目标集合为 $C=[c_1,c_2,\ldots,c_n]$,而定位则是将 IP 划分为其中的一类,如公式(6)所示:

$$\hat{c}=\arg\max_{c\in C}P(c\mid M) \tag{6}$$

假设各个测量点的测量结果是独立的,根据贝叶斯理论,可以将公式(6)表示为公式(7):

$$P(M|c)=P(\{m_1,m_2,\ldots,m_n\}|c)=P(m_1|c)\cdot P(m_2|c)\cdot\ldots\cdot P(m_n|c) \tag{7}$$

概率 $P(m_i|c)$则是通过一维核密度来进行估计的.由于测量的同时也获得了路径跳数和时延,因此两者通过加权平均来计算 $P(m_i|c)$.与待定位节点距离越近,则权值越大,将其代入可得公式(8),即定位分类器:

$$\hat{c}=\arg\max_{c\in C}(\lambda_{pop}\log(\hat{P}(c)+f_{hop}+f_{lat}) \tag{8}$$

其中, $f_{hop}=\lambda_{hop}\sum_{j=1}^{m}\exp(-j\cdot\gamma_{hop})\log(\hat{P}(h_j\mid c),f_{lat}=\sum_{j=1}^{m}\exp(-j\cdot\gamma_{hop})\log(\hat{P}(l_j\mid c)$,$\lambda_{hop},\lambda_{pop}\gamma_{hop},\gamma_{lat}$ 分别表示路径跳数、城市人口以及时延的加权系数.由公式(8)则可以估计出待定位 IP 所在的地理位置.该算法计算复杂度较小,同时具有较高精度.实验结果表明,其精度略高于 CBG 算法.

Posit 同样是对 CBG 算法的一种改进,它将所有节点分为 3 类:测量节点(M)、基准节点(N)和待定位节点(T),其中,测量节点和基准节点的地理坐标已知.首先,分别测量待定位节点与测量节点、基准节点之间的时延,通过训练获得时延与地理位置的分布函数;然后,在一个阈值λ_{lat}下,选取较小的时延计算其 L_1 范数来作为距离;最后,使用最大似然估计来预测待定位节点的位置.

Laki 等人[17]通过测量发现:时延与距离之间存在一个统一的分布规律,而且这个分布规律与基准节点的位置无关,各个分布之间彼此独立.在此基础上,提出了 Spotter 算法.在 Spotter 算法中,假设时间与地理距离之间符合标准正态分布,如公式(9)所示:

$$f_d(s)\approx\frac{1}{\sqrt{2\pi}\sigma(d)}\cdot\exp\left(-\frac{(s-\mu(d))^2}{2\sigma^2(d)}\right) \tag{9}$$

其中,$\mu(d)$,$\sigma(d)$分别表示时延分布的均值和方差,可以由参数估计的方法获得.在多个节点的情况下,时延的分布就表示为独立正态分布的联合概率.由于正态分布的参数估计方法较为简单,因此 Spotter 算法的开销较小.Arif 等人[36]采用相同思路设计了一种定位算法,所不同的是,他们用最大似然估计法来推测主机的位置,取得了更高的定位精度.

基于概率模型的定位算法主要受到概率模型的影响,因此,用于学习样本数目的多少和代表性是影响定位精度的一个重要因素.如何选择具有普遍适应性的概率模型,并准确地估计模型参数,已成为近年来定位算法的一个研究方向.

2.3 综合定位算法

上述任何一种方法都不能单独地对主机进行准确定位,为此,一些研究人员提出了综合的定位方法,典型的有 Octant[1],TTG[13],GeoWeight[34].这些定位算法同时运用上述 2~3 种算法对待定位主机进行定位,然后进行交互验证以确定主机位置.

Octant 框架[1]综合使用各种信息来对目标主机进行定位,它将所有的信息分为用正和负两类约束条件,以缩小预测区域从而提高定位精度.正约束条件是指通过主动测量来获得被测主机可能位于的区域,而负约束条件是指计算机不可能条件.它不将被测主机的位置定位于一个具体坐标,而将其可能的位置表示为一个贝赛尔曲线(Bézier curve)所决定的曲面.由于网络时延并不符合理想的时延传输模型,Octant 引入了一个“height”维来表示最后一跳的接入时延.具体说,Octant 分为如下 3 个步骤:

(1) 测量基准节点到目标节点往返时延;

(2) 归纳出强约束条件和弱约束条件,并进行排序;

© 中国科学院软件研究所 http://www.jos.org.cn

(3) 使用强约束条件进行迭代,然后使用弱约束条件进行验证.

Arif MJ 等人[34]对该算法进行了改进,进一步将预测的区域划分为更小的子区域,并为每个区域赋于不同的概率,概率越大,则节点出现的几率越高.

Yong 等人[13]发现:网络绝对时延不能精确地对主机进行定位,而只能提供较粗粒度上的定位,而许多组织网站上提供详细的地理位置,因此提出了一种 3 层定位算法(three-tiered geolocation algorithm,简称 TTG).

TTG 算法基本上分为 3 步:① 在第 1 层,使用网络时延来确定大尺度的地理位置,采用了 CBG 的改进算法;② 在第 2 层,使用大量基于网站的基准节点将 IP 定位到一个较小的范围内,通过测量点分别向基准节点和被测节点发起 traceroute 测量,通过寻找路径中的最近共同路由器来估测被测节点与基准节点间的时延,依次来确定被测主机的位置范围.由于不同的路由器到达相同被测主机的路径并不相同,TTG 算法在众多的测量结果中选择最小的时延作为两者之间的时延;③ 在第 3 层,基于第 2 层的定位信息来选择更多的通过增加基于网站基准节点的数目来选择最近基准节点以获得准确的位置(如图 4 所示).

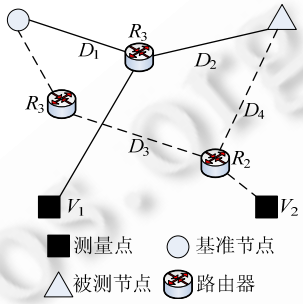


Fig.4 The principal of TTG alogrithm

图 4 TTG 算法定位原理

影响 TTG 算法精度的因素是:① 基准节点的密度;② 人口密度,人口越多,越容易获得基准节点,从而能够提高精度;③ 网络接入方式,算法精度对于最后一跳时延变化较大的网络接入方式影响较大.

总体上,综合定位算法的精度最高,计算开销最大,也是定位算法的发展方向.未来定位算法的设计需要根据用户不同定位精度的需求来进行设计,从而在计算开销与定位精度之间获得一个平衡.

3 基于客户端的定位算法

基于客户端的定位算法需要在主机或者设备上安装额外的定位装置,将定位装置提供的位置信息作为主机或设备的位置信息报告给定位系统.一般地,这些系统将 GPS、WiFi、蜂窝基站、Zigbee、射频识别(radio frequency identification,简称 RFID)、线性调频扩频技术(chirp spread spectrum,简称 CSS)等无线定位系统作为信源[35−41].

基于 GPS 的定位系统需要在主机或设备上安装 GPS 模块,主机或设备通过驱动与 GPS 模块进行交互[36].定位服务启动后,首先读取 GPS 模块的位置信息,然后根据其服务器报告其所在的位置;否则,服务器无法得知其位置所在.这种定位方式广泛地应用于移动电话、计算机以及各种嵌入式系统中,定位精度是最高的.除 GPS 外,俄罗斯的 Glonass 和我国自主研发的北斗定位系统也有一定范围的应用.

基于蜂窝、蓝牙、WiFi、RFID、Zigbee 和 CSS 的定位算法可以归纳到基于时间的系统和基于信号强度的系统两类,而且这两类系统具有一定的共性,如图 5 所示[35−40].无线定位系统一般由锚节点、待定位节点和定位服务器组成,其中,锚节点的准确地理位置已知,待定位节点为定位目标,而定位服务器则提供位置查询服务或者用户位置的解算.其中,基于蜂窝、蓝牙、WiFi、RFID 和 Zigbee 的定位算法通过测算节点之间连接信号强度(received signal strength indication,简称 RSSI),利用无线信号的空间传输衰减模型估算出节点间传输距离[42].RSSI 由于采用信号衰减进行测量,理论测量的精确距离范围在 80m 以内,80m 以外甚至将无法获得距离信

© 中国科学院软件研究所 http://www.jos.org.cn

息.实际使用时,由于环境等因素的影响,在 10m~20m 范围内定位精度可达 3m 左右.典型的基于蜂窝和 WiFi 的定位算法有谷歌公司的 My Location[43]和 Skyhook[44].

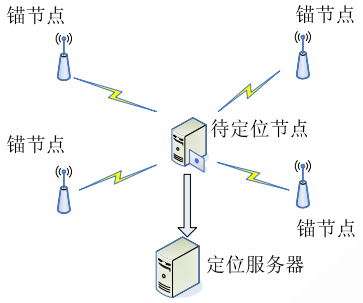


Fig.5 The principle of wireless location

图 5 无线定位原理图

CSS 采用 802.15.4a 标准,通过 SDS-TWR 的测量方法获取双向传输时延,进而获取节点距离.该算法运用窄脉冲进行精确的到达时间差(time difference of arrival,简称 TDOA)运算[45],根据定位基站的已知坐标,确定移动标签在定位场景中的位置.由于 CSS 基于时间测量机制,在测量精度 3ns~4ns 的情况下,无线电检测精度将达到 1m~1.2m.而实际使用中,由于受到前端多路径到达波检测、时间偏差等原因影响,误差可以控制在 1m~3m.

然而,这种方法存在两个问题:第一,采用这些定位方法要取得较高的精度,就必须提供尽可能多的基站等基础设施或与拥有这些设施的 ISP 进行合作,这是一般的服务提供商所无法实现的;第二,由于需要增加额外的硬件模块,会导致设备成本增加,为大面积应用带来阻力.

4 相关技术

4.1 隐私保护技术

在网络运营商和内容提供商尽量获得用户准确 IP 地址的同时,许多用户则不想暴露自己的准确地理位置而以避免隐私泄露.然而,有时只有向服务商提供地理位置才能获得用户想要的服务,这就需要设计某种机制来达到一个折中.隐私保护技术可以视为对 IP 定位技术的扩展,在用户通过独立于客户端的定位算法或基于客户端的定位算法获得准确位置后,用户使用隐私保护技术将自己的位置向定位服务器注册自己的位置.当前的隐私保护技术大致可分为两类:一类是通过伪造或者修改用户的地理位置来防止用户信息泄露[21,46−49],另一类是通过策略控制用户地理位置的可见性[50].

Kido 等人提出了一种方法,在用户的真实地址基础上生成多个虚假地址,将其发送给定位服务提供商(location based services,简称 LBS).由于这些虚假的地理位置彼此很相近,位置服务提供商也不能区别地址的真伪,从而实现对用户隐私的保护[46].另外一种常见的隐私保护方法就是采用 *k*-anonymity 方法,对用户的行踪进行伪装,这种方法既能保证用户的地理位置信息具有一定的准确性,又能保证用户隐私不被泄露,如图 6 所示.*k*-anonymity 的基本思想是:在 LBS 和用户之间设置一个受双方信任的代理,当它收到用户的 IP 地址后,它会调整用户的坐标生成一个覆盖地址,而在这个区域内至少有 $k-1$ 个其他用户.因此,在覆盖区域内的广播信息中无法确认用户的具体位置[51].同时,Beresford 等人[48]通过不断改变用户的名字来保护用户的隐私.但是在用户数据区分不大的情况下,可以通过某些方法对用户的位置进行推断[47].Liu 等人采用博弈论的方法分析了用户在使用 *k*-anonymity 方法时的行为,并对 *k*-anonymity 方法进行了分布式改进[21].

在第 2 类算法中,最简单的方法是对用户发布的位置信息采用基于组的访问控制,用户可以设置自己的朋友圈或家人看到自己位置,如 Facebook,Google+等.如今,一般的手机则只能设定用户位置可见或不可见[50],隐私控制粒度较粗.Li 等人提出一种隐私位置查询协议(privacy-preserving location query protocol,简称 PLQP),从不同粒度和层次对用户访问权限进行设定.其基本思想如下:访问粒度通过条件对用户进行过滤;将用户的地址分

© 中国科学院软件研究所 http://www.jos.org.cn

层进行加密,最高层为用户的确切地址,其余层为用户非确切地址,不同权限的用户可以对其进行解密;对协议传输过程中的所有请求和响应报文进行加密,因此,服务器也无法看到用户的确切位置.

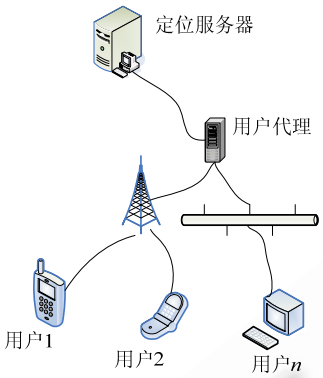


Fig.6 The main idea of *k*-anonymity

图 6 *k*-Anonymity 的基本思想

4.2 网络新技术对IP定位的影响

如今 Internet 已成为人类社会重要的基础设施,但是由于 Internet 存在 IP 地址不足、缺乏安全机制等问题,严重制约了其发展.为此,研究者提出了许多新的协议和解决方案以弥补现有 Internet 的不足.新协议改变了原有网络的结构,甚至有的协议提出了定位方法,这些新的变化对互联网 IP 定位技术产生了重大影响.

近年来,IPv6 网络基础设施建设稳步推进,商用网络部署发展迅速,用户规模成倍增长.过去两年,全球 IPv6 网络的商用部署均以超过 50%的增速在增长[52].许多用户期望 IPv6 网络中可以提供定位服务,然而 IPv6 与 IPv4 一样,不提供定位服务[53].IPv6 的地址分配中,是否将地理因素作为地址分配的因素,取决于网络运营商和互联网管理机构的决策,因此,IPv6 网络中的主机定位仍需采用额外的定位方法,而不由网络本身提供.

位置标识/身份标识分离协议(locator/ID separation protoco1,简称 LISP)是近年来提出的一种数据包路由方法[54].LISP 将网络地址分为标识符(identifier)和定位符(locator)两部分.标识符是两台主机通信会话的整个生存期内使用的位串,用于对其中一台主机相对于另一台主机进行标识.定位符是用于对某个特定包必须交付的位置进行标识的位串.在 Internet 协议中,标识符和定位符统一使用 IP 标识,这使得网络核心的路由越来越复杂.LISP 在核心路由不变的情况下,尽可能地保持了端系统的不变,用端系统的 IP 地址作为端系统的身份标识,而使用另外的路由器 ID(RLOC)作为定位符,并在路由器之间转发报文时利用这个新的 RLOC.在这种情况下,如果路由器的地理位置可以准确定位,则用以近似确定用户的地理位置.该方法的精度与 Shortest Ping 相近.

与 LISP 相近的另一种方案是主机标识协议(host identifier protocol,简称 HIP)[55,56],该协议将现有域名空间进行扩展,通过全称域名(fully qualified domain name,简称 FQDN)来命名不同的主机,以实现端系统标识符和定位符的分离.HIP 层位于传输层和网络层之间,完成主机标识与 IP 地址的双射.在 HIP 协议中,用户的地理位置可以通过域名中的信息进行推断,其定位精度和方法与 GeoTrack 类似.LISP 与 HIP 协议都可以用来对用户进行定位,但是并不能提供准确的位置,其定位精度取决于接入路由器的覆盖范围或用户提供的位置信息的精度.这主要是由于网络寻址中的定位主要是为了标识主机通信中的数据交付位置而不是物理位置所致.

HTML5 是用于取代 1999 年所制定的 HTML 4.01 和 XHTML 1.0 标准的HTML(标准通用标记语言下的一个应用)标准版本[7].在 HTML5 中,地理位置相关的主要对象是 navigator.geolocation,它有一些方法和属性检测浏览器对 HTML5 地理位置的支持性.这个 API 只定义了一个高层应用的服务接口,而不关心底层采用何种定位方法.换言之,如果用户要实现定位功能就需要在主机上安装 GPS 模块或者 WiFi 等其他硬件模块,或者采用独立于客户端的定位算法获得自己的位置,然后将定位结果提供给浏览器,才能实现定位.

© 中国科学院软件研究所 http://www.jos.org.cn

5 定位技术综合分析

前面已对各种典型的 IP 定位算法进行了详述,在本节中,我们将结合 IP 定位算法的典型评价指标对算法性能进行综合分析[57−60].

- 算法复杂度,复杂度是衡量算法计算开销的重要指标,开销越小,则算法速度越快,越易实现和部署;
- 定位精度,定位精度是衡量定位算法优劣的最重要的指标,不同粒度的定位精度,决定了算法的应用范围;
- 主动测量,主动测量会给网络产生扰动,从而影响定位精度;而被动测量或查询的方式则对网络影响较小;
- 扩展性,主要从定位系统是否存在中心化设施、部署成本以及是否容易增量部署几个方面进行衡量.分为差、中、较好 3 个层次;
- 部署方式,在定位过程中,算法的设计确定了定位系统的部署方式.采用集中式部署借助于服务器能够提高系统的收敛速度,但是不易于扩展;而分布式系统则无需基础设施的支持,具有较好的扩展性.

表 1 给出了各种定位技术的综合对比情况.

Table 1 The comprehensive comparison of different IP geolocation techniques

表 1 各种 IP 定位技术综合比较

IP 定位算法	复杂度	定位精度	客户端	主动测量	扩展性	部署方式
Geoping	$O(M(N+T))$	<50km	否	是	较好	分布
Shortest Ping	$O(MN)$	<50km	否	是	较好	分布
GeoTrack	$O(MN)$	$50<d<100$km	否	否	好	集中
CBG	$O(MN)$	$5<d<50$km	否	是	较好	集中
TBG	–	$50<d<100$km	否	是	较好	集中
Octant	$O(MN)$	35km	否	是	差	集中
Three Tier	–	690m	否	是	差	集中
Skyhook	–	<50m	否	否	较好	集中
Spotter	–	$5<d<50$km	否	是	较好	集中
GPS based	–	<50m	是	否	差	分布
Cell&WiFi based	–	<50m	是	否	较好	分布
W3C	–	–	否	–	较好	集中
EdgeScape	–	<50km	否	–	好	集中
TraceWare	–	<50km	否	–	好	集中
MaxMind	–	<50km	否	–	好	集中
IP2Geo	–	<50km	否	否	好	集中
IPligence	–	<50km	否	否	好	集中
Software77	–	<50km	否	否	好	集中
Posit	$O(M(N+T))$	43.5km	否	是	较好	集中
NBIGA	$O(MC)$	<50km	否	是	较好	分布
Geo-RX	$O(MN)$	<50km	否	是	较好	分布

其中,N 代表待定位主机,M 代表测量节点,T 代表基准节点,C 代表划分的区域数目.

从整体上看,IP 定位技术的发展趋势如下:

(1) 定位算法的设计已经由单纯的经验主义引入了更多的科学方法.最初的定位算法依靠从主机名中隐藏的信息或查询 Whois 数据库来获得主机位置,后来通过时延与地理距离来估测主机位置,在 TBG 和 Posit 等算法中,则引入了规划理论与概率估计来计算主机的位置.在未来的研究中,寻找更加合适的数学方法或建立更为准确的模型,才是提高定位精度的科学途径;

(2) 定位算法考虑了更多的网络因素,模型变得更加细致、准确.由于网路中迂回路径的存在,无法直接依靠时延与地理距离的线性关系来进行定位,因此,算法的设计过程中引入了更多的网络拓扑和路由信息,如 CBG,TBG 等算法.在未来的算法设计过程中,要考虑时延抖动、测量开销以及路径的非对称性等因素;

© 中国科学院软件研究所 http://www.jos.org.cn

(3) 从单一的定位模式向复合定位技术转变.研究表明,任意一种单一定位技术都存在其局限性,无法适用于所有的 InternetIP 设备.因此,只有各种定位方法相互验证才能提高定位精度.在设计复合定位算法或系统时应考虑定位精度的需求,采用分层设计思想,在满足用户需求的同时,尽量减少定位过程中产生的测量开销;

(4) 在提高定位精度的同时,要兼顾到用户的隐私.由于定位服务提供商尽可能地提高定位精度,而用户则希望在满足需求的情况下尽可能地隐藏自己的真实位置,未来的算法要在这两者之间达到一种平衡.

6 结束语

伴随着云计算和社交网络等新型Internet应用的不断发展,IP定位技术受到越来越多的关注,商用定位系统层出不穷,定位服务已成为 W3C 标准的一部分.研究表明:简单地利用时延与地理距离之间的线性关系和网络路径以及注册信息,任何一种技术所获得的定位精度都是有限的.综合定位方法的精度虽然较高,但是测量开销也较大,而且需要较为复杂的步骤.在设计定位算法时,不仅要追求高定位精度,同时还要充分保护用户隐私,才能取得定位服务商和用户的双赢.未来,定位算法或系统应根据不同用户的需求提供不同精度的定位服务和隐私保护,在定位复杂精度和计算开销之间达到一种平衡,并得到更多用户的支持.

References:

[1] Wong B, Stoyanov I, Sirer EG. Octant: A comprehensive framework for the geolocalization of Internet hosts. In: Proc. of the NSDI 2007 Symp. Cambridge: ACM Press, 2007. 313−326. http://www.cs.cornell.edu/People/egs/papers/octant-nsdi.pdf

[2] Azureus Software, Inc. Vuze bittorrent client. 2014. http://www.vuze.com/

[3] Katz-Bassett E, John J, Krishnamurthy A, Wetherall D, Anderson T, Chawathe Y. Towards IP geolocation using delay and topology measurements. In: Proc. of the IMC 2006. Rio de Janeiro: ACM Press, 2006. 71−84. http://homes.cs.washington.edu/~arvind/papers/geoloc.pdf [doi: 10.1145/1177080.1177090]

[4] Gill P, Ganjali Y, Dude WB. Dude, Where's that IP? Circumventing measurement-based IP geolocation. In: Proc. of the Usenix Security Symp. 2010. https://cs.uwaterloo.ca/~bernard/sec10.pdf

[5] CAIDA. CAIDA's geolocation tools comparison. 2010. http://www.caida.org/projects/cybersecurity/geolocation/

[6] IETF Geopriv Workgroup. draft-ietf-geopriv-uncertainty. 2014. http://tools.ietf.org/wg/geopriv/

[7] Robin B, Travis L, Silvia P, Erika DN, Edward O. Html5 draft. 2012. http://dev.w3.org/html5/spec-author-view/

[8] Gueye B, Uhlig S, Fdida S. Investigating the imprecision of IP block-based geolocation. In: Proc. of the PAM. 2007. 237−240. [doi: 10.1007/978-3-540-71617-4_26]

[9] IP to latitude/longitude server. 2013. http://www.iptolatlng.com/

[10] Moore D, Periakaruppan R, Donohoe J, Claffy K. Where is the world is netgeo.caida.org? In: Proc. of the INET 2000. 2000. http://www.caida.org/publications/papers/2000/inet_netgeo/inet_netgeo.html

[11] Padmanabhan VN, Subramanian L. An investigation of geographic mapping techniques for Internet hosts. In: Proc. of the ACM SIGCOMM. San Diego: ACM Press, 2001. 173−185. [doi: 10.1145/964723.383073]

[12] Gueye B, Ziviani A, Crovella M, Fdida S. Constraint-Based geolocation of Internet hosts. ACM/IEEE Trans. on Networking, 2006, 14(6):1219−1232. [doi: 10.1109/TNET.2006.886332]

[13] Wang Y, Burgener D, Flores M, Kuzmanovic A, Huang C. Towards street-level client-independent IP geolocation. In: Proc. of the 8th USENIX Conf. on Networked Systems Design and Implementation. Berkeley, 2011. https://www.usenix.org/legacy/event/nsdi11/tech/full_papers/Wang_Yong.pdf

[14] Eriksson B, Barford P, Sommersy J, Nowak R. A learning-based approach for IP geolocation. In: Proc. of the PAM. 2010. [doi: 10.1007/978-3-642-12334-4_18]

[15] Davis C, Vixie P, Goodwin T, Dickinson I. A means for expressing location information in the domain name system. 1996. http://tools.ietf.org/html/rfc1876.html

[16] Youn I, Mark BL, Richards D. Statistical geolocation of Internet hosts. In: Proc. of the Int'l Conf. on Computer Communications and Networks (ICCCN). 2009. 1−6. [doi: 10.1109/ICCCN.2009.5235373]

© 中国科学院软件研究所 http://www.jos.org.cn

[17] Laki S, Mátray P, Hága P, Sebök T, Csabai I, Vattay G. Spotter: A model based active geolocation service. In: Proc. of the INFOCOM. Shanghai: IEEE Press, 2011. 3173−3181. [doi: 10.1109/INFCOM.2011.5935165]

[18] Abboud O, Kovacevic A, Graffi K, Pussep K, Steinmetz R. Underlay awareness in P2P systems: Techniques and challenges. In: Proc. of the 23rd IEEE Int'l Symp. on Parallel and Distributed Processing (IPDPS). Rome, 2009. 1−8. [doi: 10.1109/IPDPS.2009.5160955]

[19] Cheng C, Hsiu P. Extend your journey: Introducing signal strength into location-based applications. In: Proc. of the INFOCOM. Turin: IEEE Press, 2013. 2742−2750. [doi: 10.1109/INFCOM.2013.6567083]

[20] Algis K. What is geolocation and how does it apply to network detection? 2009. http://www.sans.org/security-resources/idfaq/geolocation-network-detection.php

[21] Liu X, Liu KK, Guo L, Li X, Fang Y. A game-theoretic approach for achieving *k*-anonymity in location based services. In: Proc. of the INFOCOM. Turin: IEEE Press, 2013. 2985−2993. [doi: 10.1109/INFCOM.2013.6567110]

[22] Visualware Inc. 2014. http://www.visualroute.com

[23] NeoTrace. 2014. http://www.neoworx.com/products/neotrace/default.asp

[24] Periakaruppan R, Nemeth E. Gtrace—A Graphical Traceroute Tool. Usenix LISA, 1999.

[25] Anderson M, Bansal A, Doctor B, Hadjiyiannis G, Herringshaw G, Karplus E, Muniz D. Method and apparatus for estimating a geographic location of a networked entity. Int Cl: G06F15/16, United States Pat 6, 684, 250, 2004-1-27.

[26] MaxMind. 2012. http://www.maxmind.com

[27] EdgeScape. 2010. http://www.akamai.com/html/technology/products/edgescape.html

[28] Digital Island Inc. 2011. http://www.digitalisland.com/

[29] Software77. 2010. http://software77.net/geo-ip/

[30] IPligence. 2010. http://www.ipligence.com/products/

[31] HostIP. 2010. http://www.hostip.info/dl/index.html

[32] IPInfoDB. 2010. http://ipinfodb.com/

[33] Eriksson B, Barford P, Maggs B, Nowak R. Posit: A lightweight approach for IP geolocation. Sigmetrics Performance Evaluation Review, 2012,40(2):2−11. [doi: 10.1145/2381056.2381058]

[34] Arif MJ, Karunasekera S, Kulkarni S. GeoWeight: Internet host geolocation based on a probability model for latency measurements. In: Proc. of the ACSC 2010 33rd Australasian Conf. on Computer Science. 2010. 89−98. http://dl.acm.org/citation.cfm?id=1862209&dl=ACM&coll=DL&CFID=480394846&CFTOKEN=71466975

[35] Hu Z, Heidemann J, Pradkin Y. Towards geolocation of millions of IP addresses. In: Proc. of the IMC. 2012. 123−130. [doi: 10.1145/2398776.239879]

[36] Arif MJ, Karunasekera S, Kulkarni S, Gunatilaka A, Ristic B. Internet host geolocation using maximum likelihood estimation technique. In: Proc. of the IEEE Int'l Conf. on Advanced Information Networking and Applications (AINA). Perth: IEEE Press, 2010. 422−429. [doi: 10.1109/AINA.2010.139]

[37] Laki S, Mátray P, Hága P, Csabai I, Vattay G. A model based approach for improving router geolocation. Computer Networks, 2010,54(9):1490−1501. [doi: 10.1016/j.comnet.2009.12.004]

[38] Wang H, Wang ZY, Guobin S, Li F, Han S, Zhao F. WheelLoc: Enabling continuous location service on mobile phone for outdoor scenarios. In: Proc. of the INFOCOM. Turin: IEEE Press, 2013. 2733−2741. [doi: 10.1109/INFCOM.2013.6567082]

[39] Guha S, Plarre K, Lissner D, Mitra S, Krishna B, Dutta P, KumarS. Autowitness: Locating and tracking stolen property while tolerating gps and radio outages. In: Proc. of the SenSys 2010. 2010. 29−42. [doi: 10.1145/1869983.1869988]

[40] Zheng Y, Chen Y, Xie X, Ma WY. Geolife2.0: A location-based social networking service. In: Proc. of the MDM 2009. 2009. 357−358. [doi: 10.1109/MDM.2009.50]

[41] Constandache I, Gaonkar S, Sayler M, Choudhury RR, Cox L. Enloc: Energy-Efficient localization for mobile phones. In: Proc. of the IEEE INFOCOM Mini Conf. 2009. 2716−2720. [doi: 10.1109/INFCOM.2009.5062218]

[42] Feldmann S, Kyamakya K, Zapater A, Lue Z. An indoor Bluetooth-based positioning system: Concept, implementation and experimental evaluation. In: Proc. of the Int'l Conf. on Wireless Networks. 2003.

[43] Google maps with my location. 2009. http://www.google.com/mobile/gmm/mylocation/index.html

[44] Skyhook. 2014. http://www.skyhookwireless.com/

[45] Yan Z. Precise location technology based on chirp spread spectrum. Journal of Networks, 2011,6(6):872−878. [doi: 10.4304/jnw.6.6.872-878]

© 中国科学院软件研究所 http://www.jos.org.cn

[46] Kido H, Yanagisawa Y, Satoh T. Protection of location privacy using dummies for location-based services. In: Proc. of the 21st Int'l Conf. on Data Engineering Workshops. 2005. 1248−1252. [doi: 10.1109/ICDE.2005.269]

[47] Zang H, Bolot J. Anonymization of location data does not work: A large-scale measurement study. In: Proc. of the 17th Annualinternational Conf. on Mobile Computing and Networking. 2011. 145−156. [doi: 10.1145/2030613.2030630]

[48] Chow CY, Mokbel MF, Liu X. A peer-to-peer spatial cloaking algorithm for anonymous location-based service. In: Proc. of the ACM GIS. New York: ACM Press, 2006. 171−178. [doi: 10.1145/1183471.1183500]

[49] Beresford A, Stajano F. Location privacy in pervasive computing. IEEE Pervasive Computing, 2003,2(1):46−55. [doi: 10.1109/MPRV.2003.1186725]

[50] Li X, Jung T. Search me if you can: Privacy-Preserving location query service. In: Proc. of the INFOCOM. Turin: IEEE Press, 2013. 2760−2768. [doi: 10.1109/INFCOM.2013.6567085]

[51] Yang D, Fang X, Xue G. Truthful incentive mechanisms for *k*-anonymity location privacy. In: Proc. of the INFOCOM. Turin: IEEE Press, 2013. 2994−3002. [doi: 10.1109/INFCOM.2013.6567111]

[52] The state of the Internet. 2014. http://www.akamai.com/stateoftheinternet/

[53] Deering S, Hinden R. Internet protocol. Version 6 (IPv6) Specification. 2014. https://tools.ietf.org/html/rfc2460

[54] The locator/idenfifier separation protocol. 2008. http://www.lisp4.net/

[55] Zan FB, Xu MW, Wu JP. Survey of host identity protocol (HIP). Journal of Chinese Computer Systems, 2007,28(2):4−5 (in Chinese with English abstract). [doi: 10.3969/j.issn.1000-1220.2007.02.007]

[56] Moskwitz R, Nikander P, Henderson T. Host identity protocol. 2008. https://tools.ietf.org/html/rfc5201

[57] Muir AJ, Oorschot PC. Internet geolocation: Evasion and counterevasion. ACM Computing Surveys, 2009,42(1):1−22. [doi: 10.1145/1592451.1592455]

[58] Shavitt Y, Zilberman N. A study of geolocation databases. selected areas in communications. IEEE Journal on Networks, 2010, 29(10):2044−2056. [doi: 10.1109/JSAC.2011.111214]

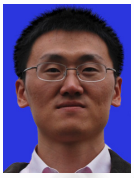
[59] Guo C, Liu Y, Shen W, Wang HJ, Yu Q, Zhang Y. Mining the Web and the Internet for accurate IP address geolocations. In: Proc. of the INFOCOM. Rio de Janeiro: IEEE Press, 2009. 2841−2845. [doi: 10.1109/INFCOM.2009.5062243]

[60] Siwpersad SS, Gueye B, Uhlig S. Assessing the geographic resolution of exhaustive tabulation for geolocating Internet hosts. In: Proc. of the PAM. Berlin: Springer-Verlag, 2008. 11−20. [doi: 10.1007/978-3-540-79232-1_2]

附中文参考文献:

[55] 昝风彪,徐明伟,吴建平.主机标识协议研究综述.小型微型计算机系统,2007,28(2):4−5. [doi: 10.3969/j.issn.1000-1220.2007.02.007]

王占丰(1982－),男,河北临城人,博士,工程师,主要研究领域为计算机网络,分布式计算,软件定义网络.
E-mail: hehengw@hotmail.com



张国敏(1979－),男,博士,讲师,主要研究领域为网络管理,分布式计算.
E-mail: zhang_gmwn@163.com

冯径(1962－),女,博士,教授,博士生导师,CCF 会员,主要研究领域为计算机网络,分布式计算,软件定义网络.
E-mail: jfeng@seu.edu.cn

许博(1980－),男,博士,讲师,主要研究领域为网络性能测量,网络流量识别,软件定义网络.
E-mail: xubo820@163.com

邢长友(1982－),男,博士,讲师,CCF 会员,主要研究领域为网络与分布式计算,未来网络,软件定义网络.
E-mail: changyouxing@126.com

© 中国科学院软件研究所 http://www.jos.org.cn