IP 定位技术研究

王志豪[1,2], 张卫东[1,2], 文 辉[1,2], 朱红松[1,2], 尹丽波[3], 孙利民[1,2]

[1] 中国科学院大学 网络空间安全学院 北京 中国 100049
[2] 中国科学院信息工程研究所 物联网安全北京市重点实验室 北京 中国 100093
[3] 国家工业信息安全发展中心 北京 中国 100040

摘要 IP 定位技术通过目标主机的 IP 地址定位其实际物理地址, 被广泛应用于定向广告、在线安全监测、网络攻击溯源等位置相关服务, 近年来实体空间资源大量接入网络空间, IP 定位受到越来越广泛的关注。本文介绍了 IP 定位的基本概念和应用场景; 根据不同应用场景分析了 IP 设备的特性; 在设备特性基础上, 对独立于设备和依赖于设备的两类定位算法进行了介绍和分析; 针对不同类型的定位技术, 介绍了 IP 定位中的攻击与防御技术; 最后对 IP 定位技术和防御技术分别进行了综合评估, 讨论了未来的发展方向。

关键词 网络定位; 网络测量; 网络攻击与防御
中图法分类号 TP309.1 **DOI 号** 10.19363/J.cnki.Cn10-1380/tn.2019.05.03

A Comprehensive Survey of IP Geolocation and Evasion

WANG Zhihao[1,2], ZHANG Weidong[1,2], WEN Hui[1,2], ZHU Hongsong[1,2], YIN Libo[3], SUN Limin[1,2]

[1] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China
[2] Beijing Key Laboratory of IOT Information Security Technology, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[3] National Industrial Information Security Development Research Center, Beijing 100040, China

Abstract IP geolocation determines geographic location by the IP address of Internet hosts. IP geolocation is widely used by target advertising, online fraud detection, cyber-attacks attribution and so on. It has gained much more attentions in these years since more and more physical devices are connected to cyberspace. In this paper, we illustrate the concept and applications of the IP geolocation. We analyze features of IP devices of different usages. Towards different IP devices, we divide IP geolocation algorithms into two categories, client-dependent and client-independent. We introduce detailed techniques of IP geolocation and evasion in this paper. At last, we compare these IP geolocation algorithms and location protection techniques.

Key words IP geolocation; network measurement; network attack and defense

1 引言

1.1 IP 定位基本概念

网络定位技术(IP 定位)是通过目标设备的 IP 地址(Internet Protocol Addresses, IP Addresses)来确定其确切的地理位置而采用的技术[1]。

从网络层开始, 网络空间的设备才能脱离物理空间的通信限制, 真正实现网络空间中的点对点互联, IP 地址是接入网络的主机在网络空间中的标识。物联网技术实现了网络空间与实体空间的融合, 人们可以通过网络来对实体空间的设备进行控制, 例如工控物联网[2], 车联网[3]。当我们将物理域接入信息域, 两者相交的部分会产生不确定性, 不确定性自然产生了安全隐患, 从而影响安全决策。地理位置信息是一项重要的实体空间信息, IP 定位则是连接网络空间标识与实体空间资源的技术, 通过 IP 定位技术, 可以将目标 IP 与其所属设备的地理位置以及其基于位置的种种价值因素联系到一起。

1.2 IP 定位技术的应用场景

位置相关的价值因素涵盖了商业价值、科研价

通讯作者: 张卫东, 硕士, Email: zhangweidong@iie.ac.cn
本课题得到国家重点研发计划(No.2016YFB0801603), 国家自然科学基金(No.61702504), 中国科学院信息工程研究所国际合作项目(No.Y7Z0451104)资助。
收稿日期: 2019-01-30; 修改日期: 2019-04-28; 定稿日期: 2019-05-13

值、安防价值等多个方面, IP 定位技术主要应用于定向内容推送、网络性能优化、社交网络、物联网、网络安全等领域。

定向内容推送。网络服务提供商根据目标的 IP 地址, 向其推送与地区相关的内容, 例如浏览器向用户推送其所在地区的新闻, 微信向用户推送附近商店等。另外, 服务提供商结合 IP 地址和其他信息追踪目标用户的常用登录地点、爱好, 推送特定的广告、服务。例如必应根据登录用户提供地区特制的搜索[4], 淘宝向用户推送特定商品等。

社交网络。社交网络根据用户的常用登录 IP 的所在地, 提供特定内容推送和潜在联系人推送, 例如, 微信会向用户推荐附近的好友。

物联网。物联网在医疗、交通、工业等领域都与位置信息紧密相关, 医疗物联网设备实时地追踪并收集地理位置信息[5, 6], 位置信息通过 LTE 网络在车联网中通信以支持各项位置相关的服务[3] 。

网络性能优化。基于 IP 定位技术提供的位置信息, 为网络节点提供更优的网络通信性能, 即更低的网络时延和资源开销。例如在 P2P 网络中为节点提供临近的对等节点从而降低远距离通信开销[7], 在车联网中通过车辆位置信息在低通信时延和能耗的前提下提高车辆间的通信距离[8]。

网络安全。网络安全问题渗透于上述各个场景之中, 因此IP定位可以应用于用户身份检验, IP伪装检测, 以及物理系统的攻击和防护。服务提供商通过用户 IP 对应的位置信息来验证用户身份从而对用户的隐私数据进行保护。IP 定位系统可以将访问者 IP 地址和其位置相关联, 检验该 IP 是否存在 IP 欺诈。IP 定位可以对网络实体系统进行定位, 从而发现网络空间中存在的实体系统[9], 同时发现系统可能存在的位置信息泄露。

1.3 本文主要贡献与结构

IP定位技术从2001年公开讨论至今已经有约18年, 然而研究者们提出的各种方法, 由于使用的实验环境(用于定位的网络主机, 采用的网络地标等等)各不相同, 对于定位技术的研究难以综合评述, 同时也缺乏从理论层面对定位技术的分析。另一方面, 国内外对于 IP 定位技术的综述工作[1, 10, 2]并不完备, 文献[1]偏重于安全方面的综述, 缺乏对定位技术的论述, 文献[10]则只是对各个定位技术进行罗列, 缺乏从定位理论上对这些技术的深入思考, 难以令人对 IP 定位技术即未来方向建立全面的认识。

本文的主要贡献如下: (1)总结现有 IP 定位技术适用的 IP 地址类型, 首次面向不同类型的 IP 设备分析了适用的 IP 定位技术; (2)分类介绍了目前为止的重要 IP 定位技术, 包括了各种 IP 定位方法以及近年来对 IP 定位理论的研究; (3)从定位逻辑上对各个定位技术进行了拆解, 首次提出了一个普适的定位理论, 并基于该理论, 讨论了 IP 定位技术未来的发展方向; (4)归纳了 IP 定位技术在安全方面的研究, 首次从定位系统和被定位用户两个角度对定位安全进行了分析。

本文第 1 节对 IP 定位技术及其应用场景进行了简单的概述。以下章节将对 IP 定位技术及其中的攻击与防御技术详细展开叙述。第 2 节对 IP 设备的特性进行归纳, 并在此基础上讨论IP定位技术的分类。第 3 节和第 4 节在第 2 节的基础上分别对两类定位技术进行细节分析。在第 5 节, 我们基于攻击与防御模型讨论了 IP 定位技术在网络安全方面的进展。我们在第 6 节对各种定位技术以及防御技术进行综合评估, 同时讨论了定位技术当前面临的挑战。最后一节对我们的工作进行了总结。

2 IP 设备与 IP 定位技术分类

IP定位技术离不开IP地址和使用该IP地址的具体设备(简称为 IP 设备)。为实现更高效的定位, 需要针对不同的IP设备的特性, 采取对应的IP定位方法, 例如: 对于智能手机, 一般来说获取 GPS 数据是最高精度的定位方法, 而对于网络路由器, 则需要通过路由器主机名等方法进行推测。因此, 我们首先需要对 IP 设备的特性进行分类, 进而根据设备的特性对 IP 定位技术进行归纳。

2.1 IP 设备的分类

由于 IP 设备的不同功能, 其在网络空间的可见性、稳定性、信息暴露程度和物理空间的位置稳定性都不尽相同。

从可见性来看, IP 地址可以分为可见地址和不可见地址, “可见”指的是 IP 设备可以响应某个网络协议, 例如 ICMP, TCP, UDP 等, “不可见”的 IP 地址则是由于通信被防火墙等机制阻断因而无法响应对应的请求。从稳定性来看, IP 地址可以分为静态地址和动态地址, 静态地址在较长时间内不会被重新分配给其他设备, 而动态地址则会频繁变更设备或是设备启用时间不固定。Heidemann 等人[11]采样检验了1%的全网地址, 他们发现 50%的样本只存活了不超过 81 分钟, 作者继而在一个更全面的普查结果中发现, 有大约 16.4%的可见 IP 地址十分稳定(在 95%的探测时间内为存活状态)。

已有研究[12-16]表明, IP 设备的定位结果往往受

到以下两个因素的影响, 一是 IP 设备的连通性, 其所处的网络环境(网络时延、拓扑等)影响了基于实时测量的定位方法, 另一个因素是设备对外暴露的信息量, 信息暴露可以分为主动公开的, 例如 WHOIS 信息详尽的 IP 段通常对应更高的位置精度。这两个因素都与设备的应用场景密切相关, 因此我们根据 IP 地址的可见性、稳定性和设备的信息暴露程度, 尽量对 IP 设备的特性进行归纳:

1) 静态 IP 设备: 这类设备的 IP 地址从网络空间上来看相对固定, 这类设备一般需要提供稳定的服务, 例如路由节点、服务器、机构网关设备等等。这些设备在实体空间来看, 也十分稳定, 其中, 服务器(网络服务器、数据库服务器、ftp 服务器、ssh 服务器)占据大约 40%[11], 这些设备往往会暴露大量信息, 其中也可能包括了地理信息, 属于活跃型设备。其他的 60%为路由器、机构网关等设备, 它们通常只响应少数协议或来源的请求, 而不会主动暴露自身信息, 因此很难挖掘其位置相关的信息, 属于被动型设备。

2) 动态 IP 设备: 网络中大部分设备的 IP 地址不固定, 例如智能手机、家庭 PC、间歇性下线的服务器等, 这些设备的地理位置的稳定性也因设备而异。有人类参与的设备通常会暴露更多的信息, 因此可以通过追踪用户的位置信息来推断 IP 地址的位置, 网站服务器已经在上一类别中讨论过, 区别只在于 IP 地址会动态变化, 因此则要求定位技术的时间效率足够高, 这类设备一般都是活跃型设备, 其中家庭 PC 和服务器的位置相对固定, 移动设备则会间歇性地变动。

3) 不可见设备: 这类设备由于长期离线或是采取了一些反探测措施, 无法通过常见的网络协议发现, 包括了离线设备、扫描阻断设备等。

2.2 IP 定位技术的分类

针对不同的 IP 设备类型, IP 定位技术主要分为两大类: 独立于客户端(client-independent)的定位技术和依赖于客户端(client-dependent)的定位技术。现今的移动设备往往携带辅助定位模块, 如 GPS、北斗、Wi-Fi、基站等, 这些辅助定位模块往往提供较高精度的位置信息, 这类技术都需要设备的硬件支持, 因此属于依赖于设备的定位技术。服务器、路由器、家庭 PC 等设备一般不包含辅助定位系统, 因此通常使用基于时延、信息推测等独立于设备的定位方法。

3 独立于客户端的 IP 定位技术

独立于客户端的 IP 定位工作在目标节点无任何辅助定位模块的状态下, 仅通过网络数据收集和网络测量等方法, 实现对目标节点的位置定位。

这类 IP 定位技术主要用到一下几个概念:

1) 探测点: 有明确位置的网络测量设备, 可以对网络中的设备发起时延、拓扑探测, 也被称为主动地标;

2) 地标: 具有明确位置的 IP 设备, 可以响应探测点的探测, 用以修正其他 IP 的位置, 通常收集自 PlanetLab 等开源地标集;

3) 目标: 待定位的 IP 设备。

Padmanabhan 等人[12]于 2001 年提出并总结了三种定位技术, GeoCluster, GeoTrack, 和 GeoPing, 分别采用三种不同的机制对 IP 进行定位, GeoCluster 认为自治系统(Autonomous System, AS)中的 IP 都存在地理上聚类的特性, 通过分析自治系统之间的边界网关协议(Border Gateway Protocol, BGP)数据, 将每个子网归属于对应的自治系统, 从而将所有子网的位置都映射到自治系统的归属机构位置, GeoTrack 方法向目标发起 traceroute, 解析路由中出现的主机名得到对应的地理名词, 并将目标 IP 位置估计为最邻近的路由节点的地理位置, GeoPing 控制探测主机通过 ping 探测目标 IP 的时延组成时延向量, 再通过与已知地标的时延向量进行对比, 计算时延向量之间的相似度, 最终将目标 IP 定位到向量距离最近的地标位置, 作者提出的三种方法为后来的许多 IP 定位方法提供了启发, 之后的许多研究工作也主要分为以下两类: 基于信息推测的 IP 定位技术和基于网络测量的 IP 定位技术。

3.1 基于信息推测的 IP 定位技术

基于信息推测的 IP 定位技术通过 IP 地址在网络中暴露的信息, 对 IP 的地理位置进行推测。由依赖的信息源不同可以分为, 基于 RDAP/WHOIS 的方法[12, 17], 基于 DNS LOC 资源记录[18]的方法, 基于主机名推断的方法[12, 19], 以及结合网络数据挖掘的方法[16, 20-21]。根据信息处理技术的难度, 可以细化为基于信息查询的定位算法和基于信息挖掘的定位算法。

3.1.1 基于信息查询

基于信息查询的定位算法, 向现有的 IP 信息源查询目标 IP 数据, 并从中得到地理信息。例如前面提到的基于 RDAP/WHOIS 的方法, 基于 DNS 的方法等等。

基于 RDAP/WHOIS 信息的定位方法, 通过直接查询网络地址注册机构(Regional Internet Registry, RIR)提供的注册者信息对 IP 地址的地理位置进行推测, 如 GeoCluster[12], NetGeo[22], WBG[23]。

基于 DNS 信息的定位方法查询 IP 地址的 DNS

记录, 从记录中的地名提示中推测地理位置, 如前面提到的 GeoTrack[12], DNS LOC 记录也可以直接得到 IP 设备的经纬度信息[18], 然而这种方法只适用于存在人工维护记录的情况。

另有一类支持穷举的特殊信息源, 即商业数据库, 例如 Cyscape[24], MaxMind's GeoIP2[25], HostIP[26], IP2Location[27], IPInfoDB[28], Software77[29]等, 商业数据库并没有公开的技术细节, 因此无法从技术角度对其进行评估, Siwpersad[30]、Shavitt[31]、Huffaker[32]以及 Gharaibeh[33]等人都通过比较定位结果对商业数据库进行了评估, 作者们一致认为: 这些数据库只是在粗粒度(国家级定位误差不低于 95%)上满足要求, 细粒度定位结果则不理想。

基于信息查询的方法比较简单, 因此易于实现, 但是问题也很明显, DNS 和 RDAP 都对地理信息和 IP 数量没有严格要求, 虽然速度快便于穷举整个网络空间, 但是定位精度较低。

3.1.2 基于信息挖掘

网络中不仅存在 IP 与地理位置"强关联"的信息, 例如 RDAP, 对任意 IP 可以找到明确的(结构化、唯一化)地理信息; 还存在大量"弱关联"的数据, 例如网页[16, 20]、地图热点(Point of Interest, POI)[16]和用户生成内容(User-generated Content, UGC)[20, 33]等, 弱关联数据意味着IP对应的地理信息是不明确的(非结构化、非唯一)。基于信息挖掘的定位算法, 指的是从网络内容中挖掘 IP 与地理位置之间的"弱关联"信息, 并从中推测可能性最大的地理位置。

Guo 等人[20]提出的 Structon 模型, 从大量网络服务器托管的网站内容中挖掘出与该服务器 IP 相关联的地理位置信息, 作者认为, 服务器托管的内容中, 可能存在与该服务器相关的地理信息。Structon 使用正则表达式在网站的网页中提取并聚合城市、国家、邮编、电话区号等字段, 生成地理信息向量, 进而基于这些已知 IP(通过 DNS 查询服务)对其他未知 IP 进行推测, 推测的依据为 IP 段的聚类规律: 作者认为前缀为/24 的子网, 所拥有的 IP 通常在同一个城市。Structon 提出的网络数据挖掘的方式提供了一个大量发现网络地标以提高定位精确度的思路。

Wang 等人[16]提出了一种精确到街道级的网络定位框架(Street-Level Geolocation, SLG), SLG 同样采用了网络服务器作为地标, 与 Structon 不同, SLG 利用地理信息服务搜索地图热点, 一些特定的地图热点包括了网站数据, 而且这些网站包含精确的地理信息, 例如学校、政府等机构网站。SLG 首先通过时延测量[13]的方法将 IP 的可能位置缩小到一个较小的区域, 然后采用如图 1 所示的相对时延测量技术, 探测节点分别向目标 IP 与地标节点发起探测, 将开始分叉的两条子路径(例如图中的 *D*1+*D*2 和 *D*3+*D*4)合起来作为相对路径, 最后将目标 IP 定位到相对路径时延最低的地标的位置。

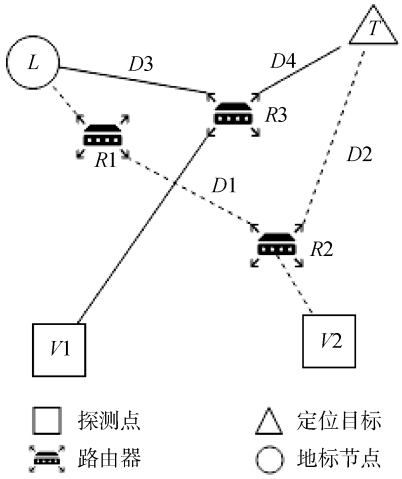

图 1 相对时延图示

Figure 1 An example of relative latency

Liu 等人[21]提出的 Checkin-Geo 基于社交网络中的用户生成内容挖掘 IP 地址的位置信息。作者发现, 用户会主动在社交网络中分享自己的签到记录, 而许多签到记录包含了位置信息, Checkin-Geo 与进而结合用户的登录日志挖掘 IP 与地理位置的对应关系。

Huffaker 等人[19]提出了一种基于 DNS 的启发式的路由器地理信息推测方法, 作者发现路由器命名方式虽然不都显式地暴露其地理位置[12](例如 corerouter1.SanFrancisco.cw.net), 但是在相同的自治域(Autonomous System, AS)中, 路由器通常以相同的规则进行命名, 比如自定义的编码格式, 因此可以从显式的地名提示出发, 挖掘域内的命名规则, 从而提高可定位路由器的比率。

Dan 等人[34]提出了基于搜索引擎日志和数据挖掘的方法, 来提高定位数据库的定位精度。作者观察到, 很多用户通过搜索引擎搜索当地的信息, 例如"北京的天气", "上海电影院关门时间"等等, 作者从搜索日志中提取用户 IP 地址并根据搜索语句中的地理名词计算位置, 最后对现有的定位数据库进行修正。

这类弱关联的数据显著地提高了基于"强关联"信息推测的 IP 精确度, 但同时也会引入噪声数据, Structon 和 SLG 可以对某些 IP 达到街道级的定位, 但是他们都是基于DNS来确定网络服务器的IP地址,

由于云服务器和 CDN 的存在, 这种方法必然会引入误差, 而文献[19]提出的地名规则方法则只适用于路由器, 而且位置精度也仅限于城市级, 为了去除噪声, 需要采用实时测量方法进行过滤。而文献[20, 33]等基于 UGC 的定位技术, 类似的方法还有 Lee 等人[35]提出的基于众包网络性能测量工具(包含 GPS 数据)的 IP 定位数据挖掘方法, 此类依赖于 UGC 的方法, 都要与对应的网络服务公司进行深入合作才能得到相关数据。

3.2 基于网络测量的 IP 定位技术

基于网络测量的 IP 定位技术旨在控制网络探测节点主动对目标进行实时探测, 根据探测数据对其进行位置估计。我们将众多测量定位技术进行了深入解析, 分为以下三个部分: 地标收集、数据采集处理和位置推断。定位服务器利用收集到的地标节点, 测量地标到目标之间的网络参数, 将收集到的数据转换为物理空间上的距离, 或直接作为距离数据进行位置推测。

数据收集处理方法和位置推断方法, 决定了定位算法的不同之处。典型的算法例如 GeoPing[12], CBG[13], TBG[14], Octant[15], SLG[16], Posit[36], MLG[37], NBIGA[38], Spotter[39], SBG[40]。

3.2.1 地标收集

地标是基于网络测量的 IP 定位技术的基础。无论是主动地标(探测点)还是被动地标, 对定位结果的影响都很大, Wang 等人[16]认为地标是定位精确度的关键。前文提到的典型算法, 它们的实验对照结果也表明, 算法的效果受到地标选取的影响很大。

常见的地标数据集, 目前主要依赖于开源数据平台、私有观测站(服务提供商提供的合作数据或私人搭建的观测站)以及网络数据挖掘方法得到的地标数据:

1) 开源数据平台提供了用于网络测量实验、应用的设备, 这些设备的提供者主要是各个学术机构, 用于研究算法、应用等在真实网络环境中的效果, 这些设备通常是一些路由器设备, 使用者也可以将这些地标作为探测点, 或是直接查看该路由器观测到的路由表。如 PlanetLab[41]、perfSONAR[43]、PingER[44], Measurement-Lab[45]等, 目前的研究文章主要都是使用了这种数据;

2) 私有地标数据是相对于开放数据平台而言的观测点或是网络地标数据, 私有观测点功能与开源实验平台相同, 只是这些信息是不公开的, 只提供给合作者私人使用。如文献[12]中采用的 Hotmail、FooTV 数据和文献[40]中使用的智能手机地标数据等;

3) 近年来提出的基于信息挖掘来获取大量的地标节点的方法, 已经在上一小节中进行了详细的介绍, 然而这些方法由于 IP 覆盖率不足, 通常需要结合网络测量对未覆盖的 IP 进行定位, 已覆盖的 IP 则作为地标。

3.2.2 数据采集与处理

在给定地标的基础上, 进行数据(时延、路由)的测量和处理: 数据测量需要控制主动地标节点通过 ping、traceroute 等工具测量得到时延、网络跳数以及数据包的传播路径, 并进行一定的误差修正, 不同算法使用的数据以及误差修正方法存在差异; 数据处理则是将测量得到的数据进行整理, 转换为算法所需要的数据格式, 取决于不同的算法, 数据处理的方式也不同, 因此我们将两个过程放在一起进行讨论。

a) 数据采集

一般来说, 数据测量方法都是使用 ping、traceroute 进行时延、拓扑探测, 直接采用往返时延的一半, 作为传播时延。一些研究者提出了进一步的改进措施, 来得到更精确的时延数据, 或是解决 ping 无法测量的问题。

Laki 等人[41]提出了一种计算单程时延的方法来提高时延测量的准确度

$$D_{pg}(s,d)=d(s,d)-H\cdot d_h$$

其中, $D_{pg}(s, d)$为单程传播时延, H 为路由跳数, d_h 为单跳时延常数。

由于一些入侵检测机制不响应 Internet Control Message Protocol(ICMP)请求, 这使得使用 ping 工具得到网络时延的方法可能失效, Muir 等人在文献[1]中提出使用 HyperText Transfer Protocol(HTTP)刷新时间来测量传输时延的方法, 作者在网站页面中加入测时脚本, 当用户访问网站时, 服务器便可以得到 HTTP 刷新时间, 从而解决了 ICMP 请求无法实现的问题。

b) 数据处理

常用的数据处理方式有, 将时延转换为实际的地理距离、将时延与路由跳数作为特征向量、将多个探测点采集到的时延组合成特征向量等。

Gueye 等人在文献[13]中使用了一种线性回归的方法将时延转换为地理距离。如图 2 所示,

1. 假设直线满足: $y=m_i x+b_i$
2. 地标测得的(g_{ij}, d_{ij})

$$y-\frac{d_{ij}-b_i}{g_{ij}}\cdot x-b_i\geqslant 0$$

3. 要求(m_i, b_i), 使得

$$\sum_{i\neq j}\left(y-\frac{d_{ij}-b_i}{g_{ij}}\cdot x-b_i\right)$$

最小。其中, y 表示往返时延(RTT), x 表示地理距离, 这样便得到了图中虚线表示的直线(bestline)。然而采用线性模型作为距离估计并不是最优的方法[14], 作者在之后又提出 GeoBuD 算法[46], 对 CBG 中的线性模型提出了修正。GeoBuD 采用分段式线性估计来消除目标与探测点之间的路由器处理延迟。GeoBuD 估算了每一跳之间的时延-距离关系

$$y_{i\tau}=m_i x_{i\tau}+b_{i\tau}$$

$$b_{i\tau}=\sum_{k=1}^{n-1}b_k$$

$$b_k=\Delta RTT_k-m_i\cdot dist(k-1,k)$$

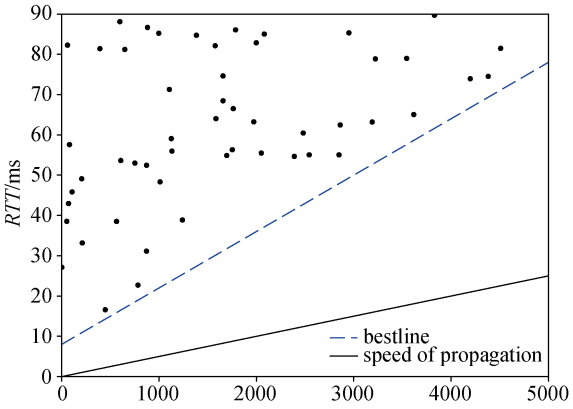

图 2 线性估计

Figure 2 Linear estimation

Katz-Bassett 等人[14]则直接将 $4c/9$ 作为时延-距离转换常数。

Laki 等人[39]基于对 PlanetLab 的时延测量数据长期观测, 发现网络时延与物理距离之间存在统计规律, 即物理距离近似符合关于时延的标准正态分布

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

其中, $\mu(d)$, $\sigma(d)$分别表示时延分布的均值和方差, 这个规律并不受探测点的影响。

Ciavarrini 等人[40]针对智能手机所处的无线环境, 统计了手机处于不同网络环境下的时延-距离规律, 从而提出了基于不同大陆和不同通信模式(3G, 4G, Wi-Fi 等等)的时延-距离转换模型。

3.2.3 位置推断

位置推断基于地标的位置, 根据处理之后的数据, 采用对应的算法, 将目标节点的位置定位绑定到地标或与地标满足某个距离约束的区域(位置绑定为其特殊情况)。根据采用的算法不同, 主要分为基于几何方法的定位技术[12-16, 40]和基于统计方法的定位技术[36-39]。

a) 基于几何方法的定位技术

基于集合方法的定位技术直接通过现有的空间理论和距离约束实现目标定位。

Gueye 等人[13]提出的基于几何约束的地理位置定位方法(Constraint-based Geolocation, CBG), CBG 改进了 GeoPing 提出的定位思路, 采用几何上的三角定位的思想(如图 3), 将目标 IP 约束在以探测点为圆心、以地理距离为半径的相交区域内。CBG 方法提供了一个清晰的定位思路, 即通过牺牲精度来快速缩小定位区域。与之前的依赖地标定位(GeoPing)的思路相比, 通过置信区间的距离约束, 将测量误差降低到置信区间的线度内, 很大程度地降低了定位的方差。

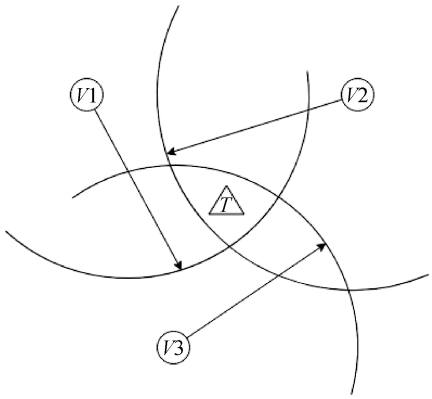

图 3 多边测量法

Figure 3 Multilateration

之后的许多算法都是在 CBG 思路的基础上, 再进一步的优化和扩展。例如, 在多边定位得到的置信区域内, CBG 直接采用区域的体心作为目标的位置, 这个估计是不严谨的, 因此, 一些算法针对置信区域, 引入了一些新的约束定位技术, 例如 Posit、SLG 等; 另一方面, 多边定位在噪声较大的网络环境(如无线环境、层级较多的网络结构)中会引入很大的误差, 因此 SBG 对无线环境下的多边定位进行了一定的流程优化。

Katz-Bassett 等人[14]提出了基于网络拓扑结构的定位方法(Topology-based Geolocation, TBG), TBG 认为需要定位目标 IP 的位置, 应该将路径中所有的中间路由器都考虑在内, 通过引入路由拓扑来进一步约束目标 IP 的地理位置, TBG 方法考虑路径中所有

相邻节点之间的两个距离约束方程:

$$d\left(V_i, x_j\right) \leqslant c_{ij}$$

和

$$d\left(x_i, x_j\right) = h_{ij} + e_{ij},$$

其中, 第一条为光速约束, 第二条为链路时延。为了降低定位误差, 作者借鉴已有的传感器网络定位算法[47], 计算使得链路总时延误差 e_{ij} 最小的坐标集合 X, 其中便包括了目标的坐标。

Wong 等人[15]提出的定位框架 Octant, 综合了时延测量、拓扑测量和主机名推测技术。从探测点发起 traceroute, Octant 计算路由中下一个节点的最优区域(计算方法类似于 CBG 的距离约束, 同时也引入了距离反约束, 即下一节点不可能位于的区域), 并将其作为另一种地标, 辅助下一个节点的定位, 直到最终到达目标。

基于几何方法的定位技术包含了一个前提, 即数据包在近似线性的空间进行传输, 然而信号传输在不同区域的速度受到许多非线性因素的影响, 例如探测点与目标之间的距离, 距离增大意味着路由跳数的增多和非传播时延的增多, 因此这种假设是过于理想化的, 定位算法需要引入更多的非线性因素。

b) 基于统计方法的定位技术

基于统计学的定位方法希望通过大量的测量数据来模拟时延/拓扑与距离的关系。

Eriksson 等人[38]提出的朴素贝叶斯分类器方法, 以 n 个探测点测得的时延和路由跃点数 $M=[m_1, m_2, \cdots, m_n]$ 作为输入数据, 目标所在郡 $C=[c_1, c_2, \cdots, c_n]$ 作为输出类别, 训练基于朴素贝叶斯理论的分类器, 最后满足

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

的郡, 即为目标的位置。

Eriksson 等人[36]提出的 Posit 与 SLG 类似, 首先使用三角定位缩小目标区域并对区域内的地标建立时延向量[12], 选择向量间 L1 范数作为训练数据, 训练, 以训练向量距离最后使用极大似然估计法推测目标的位置。

基于统计方法的定位技术, 数据密度决定了精度。文献[38]中, 作者虽然使用了 16874 个地标, 但是这些地标分布于美国全境导致地标密度很低, 因此定位误差的平均值高达 421km。

3.2.4 测量定位理论

由于测量定位技术的逐渐成熟, 对应的测量定位理论也逐渐完善。一些研究者从理论上分析了基于网络测量的 IP 定位精确度和测量定位中的安全问题。出于完整性考虑, 对测量定位安全问题的讨论将放在第 5 节, 本节主要讨论对测量定位精度有重要影响的因素(网络迂回、测量噪声和地标分布等)及其对应的研究工作。

网络迂回与中心化。Wang 等人[49]研究了网络路由在地理拓扑上的体现, 即数据包在实际物理环境中的传输路径。作者主要分析了路由的地理迂回现象和中心化现象。迂回现象指的是数据包在传输中走的实际地理路径并非地理上的最短路径, 中心化现象指的是数据包途径的关键路由往往集中于某个特定地区或国家。作者基于 CAIDA 的网络拓扑数据[50]和定位数据库, 对上述两个现象进行了观测, 作者发现迂回现象日趋严重, 尤其是较大的自治系统、同大陆以及同国家内部, 在中心化现象上, 美国掌握了大部分的关键路由节点。这些现象都对基于网络测量的定位方法产生了影响。

测量误差估计。Eriksson 等人[51]归纳 GeoPing, Octant, CBG 等方法中采用的实验环境(地标数量, 地标所处网络, 目标所处网络等等), 分析了网络地标数量和位置对定位精确度的影响。借鉴分形维度(fractal dimension), 作者引入了扩张维度(scale dimension)来刻画 IP 定位的精确度, 扩张维度描述了探测点到地标的最大时延与地标数量的关系。实验结果也验证了作者的观测。Ciavarrini 等人[52]引入了统计理论中的 Cramér-Rao 下限(Cramér-Rao lower bound, CRLB)来评估时延-距离估计的最小均方误差, CRLB 可以有效地为测量定位方法的精确度提供理论上线。与文献[46]类似, 作者基于实验结果得出, 地标密度影响了测量定位的精确度。

3.3 小结

独立于客户端的 IP 定位方法, 从技术上可以拆解为信息采集、信息聚合、距离计算和位置推测四个过程。信息采集指的是从网络中搜索目标的各方面信息(格式化或非格式化), 不同数据源或多或少地会暴露其位置相关的数据。从不同数据源获取的地理信息存在重复和差异, 通过信息聚合可以直接得到一部分地标, 这类地标包括了开放的网络探测平台以及许多位置明确的网络服务器、路由器等, 在第 2.1 节我们已经讨论过, 稳定 IP 设备大概只占全网地址的 16.6%, 其中网络服务器、路由器可以通过一些方法[12, 15, 19]进行定位, 大部分 IP 还是需要基于地标进行距离估计和位置推测。距离估计方法并不限于基于时延的线性/非线性拟合[13-15, 39], 最长前缀匹配[12, 20]、时延向量距离[12, 36]、相对时延[16]等方式从广义上来说都属于距离

估计。最后在地标和距离的基础上, 可以选择多种位置估计方法, 例如许多技术使用的就近原则[12, 16, 23, 53]、三角定位方法[13, 15, 16, 36, 46]、极大似然估计法[37]、朴素贝叶斯分类[38], 对目标进行位置估计。

当前的 IP 定位技术通常会综合多项技术, 从前面的技术分析可以看出, 定位精度和准确度主要依赖于目标区域的地标分布密度和准确度(取决于IP相关的信息量)和目标所处的网络环境, 而且基于测量的定位技术都存在同样的问题, 即很难实现实时的计算。近些年, 由于测量定位技术的逐渐成熟, 测量定位方面的主要研究重心也从对具体定位技术的研究变为对测量定位理论和测量定位安全的研究。

4 依赖于客户端的 IP 定位技术

依赖于客户端的定位技术, 顾名思义, 需要被定位的目标有额外的定位装置, 定位服务器向目标申请定位装置的访问权限, 进而基于定位装置进行定位。由于此类 IP 定位, 其结果精度主要依赖于具体的被定位设备, 因此与本文的主体相关度不大, 在此处只进行简单介绍。根据定位装置与测量点之间的距离, 可以分为远程定位技术和局域定位技术。

4.1 远程定位技术

远程定位技术主要指的是卫星定位, 例如 GPS[54]、北斗系统[55]。设备中的 GPS 模块与卫星定位系统实时通信从而计算出自身的地理位置, 在理想条件下, 卫星定位的精度可以达到米级。

卫星定位技术与上一节中的网络测量定位虽然都属于远程定位技术, 但是卫星定位的精度相对于网络测量定位高了 3-4 个量级。主要原因在于, 网络数据包从探测点到目标的传输过程中受到大量不可测的误差(传输时延、排队时延等), 误差的数量级甚至高于测量数据(传播时延)的数量级, 而卫星定位主要受到卫星端的轨道和钟差、电流层和对流层延迟、接收端的测量噪声和多路径效应的影响, 干扰量级较小, 而且卫星定位存在探测点与目标之间的信息交互, 可以有效地对误差进行控制。

卫星定位的不足之处在于建筑物的遮挡导致室内定位的效果很差, 近年来有一些研究致力于提高 GPS 在室内的定位效果[56, 57]。

4.2 局域定位技术

局域定位技术通过目标附近的测量点对目标进行实时测量, 然后根据信号特性对客户端进行定位。定位的信道取决于客户端搭载的信号模块, 例如 Wi-Fi、蓝牙、蜂窝网、RFID 等等[58-63]。按照数据类型, 局域定位技术可以分为基于到达时间(time of arrival, TOA)、到达时间差(time different of arrival, TDOA)、到达角度(angle of arrival, AOA)和接收信号强度(received signal strength indication, RSSI)的定位方法; 按照位置推测方法, 可以分为基于距离的定位方法, 和基于指纹的方法。

基于距离的方法将 TOA/TDOA/AOA/RSSI 转换为探测点到目标的实际距离, 然后使用几何方法进行定位。

基于指纹的方法, 根据多探测点测得的数据生成位置相关的强度向量, 对目标的强度向量进行相似度匹配来实现定位。

4.3 小结

依赖于客户端的定位方法, 无论是远程定位还是局域定位方法, 探测点与目标之间都是直接通信, 可以直接测得客户端的数据, 因此定位精度远高于独立于客户端的定位方法。然而此类定位方法的缺点也很明显, 一方面是这类方法只适用于有特殊定位模块的设备(例如智能手机), 另一方面, 测量点的控制权限会受到通信服务提供商的限制。

5 IP 定位中的攻击与防御

IP 定位的目的是解决三个不同标识(IP、用户和地理位置)之间的关联问题[1]:

1. 终端用户的定位: 确定目标用户的地理位置;
2. 目标 IP 的定位: 确定目标 IP 的地理位置;
3. 用户 IP 的识别: 确定用户使用的 IP。

从攻击者(定位者)和防御者(被定位者), 可以分以下两个角度进行讨论。

5.1 IP 定位攻击

从攻击者角度, 被攻击者(被定位者)可以分为两种: (1)被定位者暴露真实 IP; (2)被定位者隐藏真实 IP。两种情况下, 定位者都希望得到被定位者的真实地理位置。

被定位者(1)所处情况为最一般的情况, 此时问题 3 是已解决的, 这种情况下问题 1 等价于问题 2。攻击者只需要通过第 3 节和第 4 节中介绍的技术实现位置定位。

被定位者(2)通过分离自己与公网 IP 的方法实现隐藏, 常见的隐蔽方法有拨号上网, 代理上网和远程会话:

1. 远距离拨号上网。用户通过电话线拨号上网[64]可以连接远距离的 ISP[65, 66], 从而分离 IP 与用户自身的地理位置;

2. 代理上网。用户将网络请求通过匿名网络(如

Tor[67], Privoxy[68])转发到目标地址, 从而隐藏自己的真实 IP;

3. 远程会话。通过远程登录, 控制远程计算机访问网络, 例如 Windows 远程桌面, X11 转发[69], VNC(Virtual Network Computing)[70]等等。远程会话与方法 2 的本质区别在于应用程序运行在何处, 方法 2 只借助远程机转发网络连接, 应用程序依然在本地机, 而方法 3 直接在远程机上运行应用程序。

通过远程访问网络隐藏 IP 的用户, 无论从应用层以下都脱离了本地机器, 只能定位到用户登录的远程机, 因此没有一个可行的用户定位方法。通过远距离拨号和代理上网, 虽然只是隐藏了 IP 与用户之间的关系, 但是依然可以通过依赖于客户端的定位方法实现对用户的定位。Weinberg 等人[71]发现, 攻击者可以通过时延探测定位代理服务器的位置。另一方面, 攻击者可以利用一些客户端的漏洞, 发现被定位者隐藏的 IP, 例如 Muir 等人[1]发现某些版本的 Java 虚拟机可以绕过代理。

5.2 IP 定位防御

从防御者角度, 假定攻击者(定位者)通过定位系统确定防御者(被定位者)的位置。攻击者可以直接通过防御者的定位模块获取位置信息, 或者通过探测点和地标对防御者进行探测; 防御者可以控制自己的设备伪造位置信息, 同时可以控制攻击者与被攻击者之间的物理链路。针对攻击者采用的两类定位方法(依赖于客户端的和独立于客户端的), 防御者可以采用两类防御技术: 信息伪造和测量数据伪造。

5.2.1 信息伪造

信息伪造技术针对攻击者采用的依赖于客户端的定位方法和基于信息推测的定位方法, 通过混淆化、匿名化和隐私策略, 向攻击者提供伪造的信息, 从而实现对 IP 定位攻击的防御。

混淆化。Kido 等人[72]伪造多个假地址, 与真实地址一起发送给攻击者, 由于假地址与真实地址相近, 攻击者很难分辨攻击者的真实位置。

匿名化。匿名化指的是防御者通过可信代理访问连接攻击者(定位服务器), 可信代理在转发攻击者位置信息时去除攻击者 IP, 并采用一定的隐蔽算法修改防御者的地理位置[73-76]。常见的匿名化方法为 k-anonymity, 即代理服务器可以提供至少 k 个接入设备的匿名化。Beresford 等人[74]在匿名组内通过频繁伪造身份标识进一步增强匿名化。由于 k-anonymity 需要引入额外的能耗进行身份和地址的伪造, Liu 等人[77]从博弈的角度分析了基于 k-anonymity 的算法能耗, 提出了基于博弈论的策略优化方法。

隐私策略, 即控制位置隐私的访问权限。Li 等人[78]提出了一种细粒度的隐私保护协议(Privacy-preserving Location Query Protocol, PLQP), 根据访问者的身份防御者可以限制其观察到的位置信息。

5.2.2 测量数据伪造

测量数据伪造技术针对攻击者采用的基于网络测量的定位方法, 通过控制协议栈或临近的网络路由器实时构造响应包, 向攻击者提供伪造的测量数据, 从而实现对 IP 定位的防御。

Gill 等人[79]提出了针对基于时延探测的定位系统的防御方法, 作者通过修改时延和网络拓扑, 在一定程度上影响了定位系统的判断。

时延伪造。Gill 等人[79]通过推迟响应, 以提高探测点测得的网络时延, 达到了误导定位的效果, 然而这个方法很难精确地误导到某指定地区, 因此这种操作可能会产生很大的定位误差从而被定位系统发现。Abdou 通过修改 ICMP 响应包和提前发送 ICMP 不可达信息[80], 达到增加和减小探测时延的效果, 从而影响定位。

拓扑伪造。文献[79]同时提出了面向网络拓扑的防御方法, 作者控制防御者周围的路由器将探测包导向其他网关路由器, 达到修改路由的效果。

5.3 小结

由于涉及位置隐私, IP 定位的安全问题受到了广泛的关注, 然而研究者通常只关注基于客户端的位置隐私保护[72-78], 较少提及独立于客户端的定位安全问题中[1, 79-81]。从定位者角度, 定位算法通常以被定位者不会修改自身状态为前提进行定位, 然而一旦被定位者隐藏自己的 IP, 独立于客户端的定位算法几乎全部无效, 基于客户端的定位算法只能在被定位者允许下实现用户定位, 无法直接关联 IP。从被定位者角度, 隐蔽 IP、伪造位置信息、伪造时延/拓扑信息都可以成功达到位置保护的功能, 然而这类方法都需要付出额外的代价: 拨号上网的方法由于连接速度的限制很少被使用; 代理服务器在某些情况下可以被定位; 信息伪造往往只适用于移动端的位置保护, 而且需要额外的能量损耗; 非移动端设备虽然可以通过时延构造的方法达到位置伪造, 但是需要对其周围的网络有控制权, 这对于个人用户是不可能实现的。

6 IP 定位技术的综合评估

IP 定位技术评估。前文已经详细介绍了 IP 定位算法及其安全研究的现状, 我们在定位算法的计算复杂度、部署复杂度、定位精度对定位算法的精度进行比较。

表 1 独立于客户端的定位算法性能比较

Table 1 Comparison of client-independent algorithms

算法	单目标复杂度	地标部署	定位误差
GeoPing	$O_p(VL)$	14(美国)	~400km (med.)
GeoTrack	$O_t(V)$	14(美国)	~590km (med.)
CBG	$O_p(V)$	42(欧洲)	~22km (med.)
		95(美国)	~95km (med.)
TBG	$O_p(V)$	11(美国)	~200km (med.)
		68(美国)	~180km (med.)
		128(美国)	~67km (med.)
Octant	$O_t(V)+O_p(V)$	104(美国)	~35km (med.)
NBIGA	$O_p(VL)$	225(美国)	~200km (med.)
Posit	$O_p(VL)$	431(美国)	~44km (med.)
SLG	$O_p(V)+O_t(L)$	~76000(美国)	~2km (med.)

其中 $O_t(1)$ 和 $O_p(1)$ 分别表示单次 traceroute 的时间复杂度和单次 ping 的时间复杂度, V 表示探测点数量, L 表示地标数量。算法的部署复杂度通过部署探测点和地标的数量来体现, 由于依赖客户端定位模块的定位方法的局限性通常在于用户所处环境, 难以纵向比较, 表 1 只比较了独立于客户端的定位算法。由表 1 可以看出:

1. 基于网络测量的定位技术, 在不同地区的定位结果差别很大, 其精确度通常取决于网络地标和网络环境。例如, CBG 在欧洲的定位结果通常优于在美国的定位结果, 因为其使用的 PlanetLab 地标在欧洲的密度远高于美国; TBG 采用同样的 PlanetLab 地标(教育网环境)对不同目标网络的定位结果差异很大(在教育网环境内约 67km, 而非教育网环境则有大约 200km);

2. 基于信息挖掘的定位技术, 可以对一部分 IP 地址直接实现精确定位, 如 Structn 和 SLG 方法, 此类方法的局限性在于 IP 地址的覆盖和定位结果的时效性。信息挖掘定位只能覆盖部分有用户使用或是具有独特信息的 IP 设备(如网络服务器), 难以覆盖其他 IP, 因此通常需要采用其他方法提高覆盖率, 同时也引入了误差; 同时, 虽然此类方法可以确定 IP 设备在一段时间内的位置信息, 但是由于网络的动态变化, 尤其是一般用户 IP 和网络服务器 IP, 此类方法的时效性往往不高;

3. 从单纯测量的技术来说, Octant 达到了最好的定位结果, 但是相比于其他方法(GeoTrack, CBG 等), 大量拓扑探测也意味着更高的时间开销; 大的地标密度可以大幅提高定位精确度, 但是也意味着非常高的时间复杂度和部署/发现网络地标的难度;

4. 首先基于多边定位缩小置信区域, 进而在置信区域内大量地标的细粒度定位方法(Posit 和 SLG)可以达到较高的定位精度, 也是目前的一个重要的研究方向;

5. 现有信息挖掘方法主要针对用户 IP、服务器 IP 和路由器 IP, 其他细分 IP 设备类型的信息挖掘方法也是一个重要的研究方向。

IP 定位安全评估。进一步的, 我们评估了针对 IP 定位方法的防御技术。根据我们在第五节开头提出的三个目的, 定位防御技术的最终目标是防止定位算法得到用户的真实地理位置, 或者分离用户的表象 IP 与其入网的真实 IP。

表 2 IP 定位防御技术评估

Table 2 Comparison of defense methods against IP geolocation algorithms

	独立于客户端		依赖于客户端
	信息推测	网络测量	
远程拨号上网	可规避	可规避	不可规避
代理	可规避	可规避	可规避 IP 定位 不可规避用户定位
远程会话	可规避	可规避	可规避
位置信息伪造	不可规避	不可规避	可规避
测量数据伪造	不可规避	可规避	不可规避

从上表可以看出, IP 隐蔽技术(分离用户的真是 IP 与表象 IP)相对于位置隐蔽技术可以更好地规避 IP 定位攻击。当 IP 隐蔽失效时, 位置信息混淆可以有效地防护基于客户端的定位技术, 无客户端支持的定位技术则需要通过测量数据伪造的方法进行规避, 这两种方法都无法保护基于 IP 信息推测的定位攻击。

7 结论

我们在本文总结了 IP 定位技术, 并提出了 IP 定位中的攻击与防御模型, 并在此模型基础上, 总结了当前的位置隐私保护技术。

由于远距离的距离转换算法效果无法令人满意, 因而当前的 IP 定位技术, 逐渐往细粒度、局域的高

地标密度发展, 同时也需要更高的探测时间复杂度和数据积累的难度。因此, 在时间效率与定位精确度之间存在平衡, 细粒度的位置推测和低时间响应算法还需要深入研究。

在防御 IP 定位攻击方面, 以往的算法往往只关注于移动用户的位置保护, 而对于家庭 PC、网络服务器、路由器等无定位模块的主机, 仅靠 IP 隐蔽依然可能存在问题, 应进一步研究基于测量数据伪造的防御技术, 以达到更全面的保护。

参考文献

[1] J. A. Muir and P. C. V. Oorschot, “Internet geolocation: evasion and counter evasion,” in *ACM Computing Surveys (CSUR’09)*, vol. 42, no. 1, p. 4, 2009.

[2] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, “The industrial internet of things (IIoT): An analysis framework,” in *Computers in Industry*, vol. 101, pp. 1-12, 2018.

[3] H. Seo, K. Lee, S. Yasukawa, Y. Peng, and P. Sartori, “LTE evolution for vehicle-to-everything services,” in *IEEE Communications Magazine*, vol. 54, no. 6, pp. 22-28, June 2016.

[4] “Bing,” Microsoft, https://cn.bing.com/.

[5] C. Costa, C. Pasluosta, B. Eskofier, D. Silva, and R. Righi, “Internet of Health Things: Toward intelligent vital signs monitoring in hospital wards,” in *Artificial Intelligence in Medicine*, vol. 89, pp. 61-69, 2018.

[6] A. Engineer, E.M. Sternberg, and B. Najafi, “Designing Interiors to Mitigate Physical and Cognitive Deficits Related to Aging and to Promote Longevity in Older Adults: A Review,” in *Gerontology*, 2018.

[7] O. Abboud, A. Kovacevic, K. Graffi, K. Pussep, and R. Steinmetz, “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.

[8] F. Martin-Vega, B. Soret, M. Torres, I. Kovács, and G. Gómez, “Geolocation-Based Access for Vehicular Communications: Analysis and Optimization via Stochastic Geometry,” in *IEEE Trans. Vehicular Technology*, vol. 67, no. 4, pp. 3069–3084, 2018.

[9] Q. Li, X. Feng, H. Wang, and L. Sun, “Understanding the Usage of Industrial Control System Devices on the Internet,” in *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2178-2189, June 2018.

[10] Z.F. Wang, J. Feng, C.Y. Xing, G.M. Zhang, and B. Xu, “Research on the IP Geolocation Technology,” in *Journal of Software*, 2014, 25(7): 1527-1540.
(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。)

[11] J. Heidemann, Y. Pradkin, R. Govindan, C. Papadopoulos, G. Bartlett, and J. Bannister, “Census and Survey of the Visible Internet”, in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement (IMC’08)*, pp. 169-182, 2008.

[12] V. N. Padmanabhan and L. Subramanian, “An investigation of geographic mapping techniques for Internet hosts,” in *ACM SIGCOMM Computer Communication Review (CCR’01)*, vol. 31, no. 4, pp. 173–185, 2001.

[13] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida, “Constraint-based geolocation of Internet hosts,” in *IEEE/ACM Transactions On Networking (TON’06)*, vol. 14, no. 6, pp. 1219–1232, 2006.

[14] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Ander-son, and Y. Chawathe, “Towards IP geolocation using delay and topology measurements,” in *Proc. of the 6th ACM SIGCOMM conference on Internet measurement (IMC’06)*, pp. 71–84, 2006.

[15] B. Wong, I. Stoyanov, and E. G. Sirer, “Octant: A comprehensive framework for the geolocalization of Internet hosts,” in *Proc. of the 4th USENIX Symposium on Networked Systems Design & Implementation (NSDI’07)*. USENIX Association, 2007.

[16] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang, “Towards Street-level Client-independent IP Geolocation,” in *NSDI*, 2011.

[17] “Registration Data Access Protocol (RDAP) Query Format,” RFC-7482, IETF, https://tools.ietf.org/html/rfc7482, Mar. 2015.

[18] “A Means for Expressing Location Information in the Domain Name System,” RFC-1876, IETF, https://tools.ietf.org/html/rfc1876, Jan. 1996.

[19] B. Huffaker, M. Fomenkov, and k. claffy, “DRoP:DNS-based Router Positioning,” in *ACM SIGCOMM Computer Communication Review (CCR)*, vol. 44, no. 3, pp. 6-13, Jul 2014.

[20] C. Guo, Y. Liu, W. Shen, H. J. Wang, Q. Yu, and Y. Zhang, “Mining the web and the Internet for accurate IP address geolocations,” in *Proc. of IEEE Conference on Computer Communications (INFOCOM’09),* pp. 2841–2845, 2009.

[21] H. Liu, Y. Zhang, Y. Zhou, D. Zhang, X. Fu and K. K. Ramakrishnan, “Mining checkins from location-sharing services for client-independent IP geolocation,” in *Proc. of IEEE Conference on Computer Communications (INFOCOM’14)*, pp. 619-627, 2014.

[22] D. Moore, R. Periakaruppan, J. Donohoe, and K. Claffy, “Where in the world is netgeo.caida.org?” in *International Networking Conference (INET’00)*, Jul. 2000.

[23] P. T. Endo and D. F. H. Sadok, “Whois Based Geolocation: A Strategy to Geolocate Internet Hosts,” in *24th IEEE International Conference on Advanced Information Networking and Applications*, pp. 408-413, Perth, WA, 2010.

[24] “Cyscape’s CountryHawk,” http://www.cyscape.com/products/chawk/.

[25] “MaxMind’s GeoIP2,” http://www.maxmind.com.

[26] “HostIP,” http://www.hostip.info/dl/index.html.

[27] “IP2location,” http://www.ip2location.com/.

[28] “IPInfoDB,” http://ipinfodb.com/.

[29] “Software77,” http://software77.net/geo-ip/.

[30] S. S. Siwpersad, B. Gueye, and S. Uhlig, “Assessing the geographic resolution of exhaustive tabulation for geolocating internet hosts,” in *Proc. of PAM*, 2008.

[31] Y. Shavitt and N. Zilberman, “A Study of Geolocation Databases,” *ArXiv e-prints*, May 2010.

[32] B. Huffaker, M. Fomenkov, and K. Claffy, “Geocompare: A Comparison of Public and Commercial Geolocation Databases,” *Technique report*, 2011.

[33] M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, and C. Papadopoulos, “A Look at Router Geolocation in Public and Commercial Databases,” in *Internet Measurement Conference (IMC)*, Nov 2017.

[34] O. Dan, V. Parikh, and B. D. Davison, “Improving IP Geolocation using Query Logs,” In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM '16)*, pp. 347-356, 2016.

[35] Y. Lee, H. Park, and Y. Lee, “IP Geolocation with a Crowd-sourcing Broadband Performance Tool,” in *SIGCOMM Comput. Commun. Rev.*, vol. 46, no. 1, pp. 12-20, Jan. 2016.

[36] B. Eriksson, P. Barford, B. Maggs, and R. Nowak, “Posit: a lightweight approach for IP geolocation,” in *SIGMETRICS Perform. Eval. Rev.*, vol. 40, pp. 2-11, 2012.

[37] I. Youn, B. L. Mark, and D. Richards, “Statistical geolocation of Internet hosts,” in *Proc. of 18th IEEE International Conference on Computer Communications and Networks (ICCCN’09)*, pp. 1–6, 2009.

[38] B. Eriksson, P. Barford, J. Sommers, and R. Nowak, “A learning-based approach for IP geolocation,” in *Proc. of International Conference on Passive and Active Network Measurement (PAM’10)*, pp. 171–180, 2010.

[39] S. Laki, P. M´atray, P. H´aga, T. Seb˝ok, I. Csabai, and G. Vattay, “Spotter: A model based active geolocation service,” in *Proc. of IEEE Conference on Computer Communications (INFOCOM’11)*, pp. 3173–3181, 2011.

[40] G. Ciavarrini, V. Luconi, and A. Vecchio, “Smartphone-based geolocation of Internet hosts,” in *Computer Networks*, vol. 116, pp. 22-32, 2017.

[41] S. Laki, P. M´atray, P. H´aga, I. Csabai, and G. Vattay, “A model based approach for improving router geolocation,” in *Computer Network*, pp. 1490–1501, 2010.

[42] “PlanetLab,” https://www.planet-lab.org.

[43] A. Hanemann, J. W. Boote, E. L. Boyd, J. Durand, L. Kudarimoti, R. Łapacz, D. M. Swany, S. Trocha, and J. Zurawski, “PerfSONAR: A Service Oriented Architecture for Multi-domain Network Monitoring,” in *International Conference on Service-Oriented Computing*, pp. 241-254, 2005.

[44] W. Matthews and L. Cottrell, “The PingER project: active Internet performance monitoring for the HENP community,” in *IEEE Communications Magazine*, vol. 38, no. 5, 2000.

[45] “Measurement-Lab,” https://www.measurementlab.net.

[46] B. Gueye, S. Uhlig, A. Ziviani, and S. Fdida, “Leveraging Buffering Delay Estimation for Geolocation of Internet Hosts,” in *International Conference on Research in Networking*, 2006.

[47] P. Biswas and Y. Ye, “Semidefinite programming for ad hoc wireless sensor network localization,” in *Proceedings of Information Processing in Sensor Networks*, 2004.

[48] L. Doherty, K. S. J. Pister, and L. E. Ghaoui, “Convex position estimation in wireless sensor networks,” in *Proceedings of Infocom*, pp 1655–1633, 2001.

[49] J. H. Wang and C. An, “A study on geographic properties of internet routing,” in *Computer Networks*, vol. 133, pp. 183-194, 2018.

[50] CAIDA, “The IPv4 Routed /24 Topologytaset,” https://www.caida.org/data/active/ipv4_routed_24_topology_dataset.xml.

[51] B. Eriksson and M. Crovella, “Understanding geolocation accuracy using network geometry,” in *Proceedings IEEE INFOCOM*, pp. 75-79, 2013.

[52] G. Ciavarrini, M. S. Greco, and A. Vecchio, “Geolocation of Internet hosts: Accuracy limits through Cramér–Rao lower bound,” in *Computer Networks*, vol. 135, pp. 70-80, 2018.

[53] D. Li, J. Chen, C. Guo, Y. Liu, J. Zhang, Z. Zhang, and Y. Zhang, “IP-Geolocation Mapping for Moderately Connected Internet Regions,” in *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 2, pp. 381-391, Feb. 2013.

[54] “GPS,” https://www.gps.gov/.

[55] “北斗卫星导航系统”, http://www.beidou.gov.cn/.

[56] “羲和北斗”, http://www.xhbds.cn/.

[57] J. Barnes, C. Rizos, J. Wang, D. Small, G. Voigt, and N. Gambale, “Locata: A New Positioning Technology for High Precision Indoor and Outdoor Positioning,” in *Proceedings 2003 International Symposium on GPS\GNSS*, 2003.

[58] S. Feldmann, K. Kyamakya, A. Zapater, and Z. Lue, “An indoor Bluetooth-based positioning system: Concept, implementation and experimental evaluation,” in *Proc. of the Int’l Conf. on Wireless Networks*, 2003.

[59] H. Wang, Z. Wang, G. Shen, F. Li, S. Han, and F. Zhao, “WheelLoc: Enabling continuous location service on mobile phone for outdoor scenarios,” in *Proceedings IEEE INFOCOM*, pp. 2733-2741, Turin, 2013.

[60] I. Constandache, S. Gaonkar, M. Sayler, R. R. Choudhury, and L. Cox, “EnLoc: Energy-Efficient Localization for Mobile Phones,”

in *IEEE INFOCOM 2009*, Rio de Janeiro, pp. 2716-2720, 2009.

[61] S. Guha, K. Plarre, D. Lissner, S. Mitra, B. Krishna, P. Dutta, and S. Kumar, "AutoWitness: locating and tracking stolen property while tolerating GPS and radio outages," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems (SenSys '10)*, 2010.

[62] A. Makki, A. Siddig, M. Saad, and C. Bleakley, "Survey of WiFi positioning using time-based techniques," in *Computer Networks*, vol. 88, pp. 218-233, 2015.

[63] A. Günther and C. Hoene, "Measuring Round Trip Times to Determine the Distance Between WLAN Nodes," in *International Conference on Research in Networking*, 2005.

[64] M. Hauben and R. Hauben, "Netizens: On the History and Impact of Usenet and the Internet (1st ed.)," Los Alamitos, CA: IEEE Computer Society Press, pp. 161–200, 1997.

[65] "PemTel," https://www.pemtel.com/dial-up-internet.html.

[66] "IgLou," https://www.iglou.com/local-access/.

[67] "Tor," https://tor.eff.org/about/overview.html.en.

[68] "Privoxy," http://www.privoxy.org/.

[69] "X11 forwarding," http://en.tldp.org/HOWTO/XDMCP-HOWTO/ssh.html.

[70] T. Richardson, Q. Stafford-Fraser, K. R. Wood, and A. Hopper, "Virtual network computing," in *IEEE Internet Computing*, vol. 2, no. 1, pp. 33-38, Jan.-Feb. 1998.

[71] Z. Weinberg, S. Cho, N. Christin, V. Sekar, and P. Gill, "How to Catch when Proxies Lie: Verifying the Physical Locations of Network Proxies with Active Geolocation," in *Proceedings of the Internet Measurement Conference 2018 (IMC '18)*, 2018.

[72] H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of Location Privacy using Dummies for Location-based Services," in *21st International Conference on Data Engineering Workshops (ICDEW'05)*, Tokyo, Japan, pp. 1248-1248, 2005.

[73] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. ACM MobiSys*, pp. 31–42, 2003.

[74] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," in *IEEE Pervasive Computing*, vol. 2, no. 1, pp. 46-55, Jan.-March 2003.

[75] H. Zang and J. Bolot, "Anonymization of location data does not work: a large-scale measurement study," in *Proceedings of the 17th annual international conference on Mobile computing and networking (MobiCom '11)*, ACM, New York, NY, USA, 145-156, 2011.

[76] C. Chow, M. F. Mokbel, and X. Liu. "A peer-to-peer spatial cloaking algorithm for anonymous location-based service," in *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems (GIS '06)*, ACM, New York, NY, USA, 171-178, 2006.

[77] X. Liu, K. Liu, L. Guo, X. Li, and Y. Fang, "A game-theoretic approach for achieving k-anonymity in Location Based Services," in *2013 Proceedings IEEE INFOCOM*, pp. 2985-2993, 2013.

[78] X. Li and T. Jung, "Search me if you can: Privacy-preserving location query service," in *2013 Proceedings IEEE INFOCOM*, pp. 2760-2768, 2013.

[79] P. Gill, Y. Ganjali, B. Wong, and D. Lie, "Dude, where's that IP?: circumventing measurement-based IP geolocation," in *Proceedings of the 19th USENIX conference on Security (USENIX Security'10)*, USENIX Association, Berkeley, CA, USA, 16-16, 2010.

[80] A. Abdou, "Internet Location Verification: Challenges and Solutions," *ArXiv e-prints*, 2018.

[81] C. Castelluccia, M. A. Kaafar, P. Manils, and D. Perito, "Geolocalization of proxied services and its application to fast-flux hidden servers," in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement (IMC '09)*, 2009.

王志豪 于 2015 年在中国科学技术大学理论物理专业获得理学学士学位。现在中国科学院大学信息工程研究所通信与信息系统专业攻读博士学位。研究领域为网络定位、物联网安全。研究兴趣包括: 网络定位、安全大数据分析。Email: wangzhihao@iie.ac.cn

张卫东 于 2012 年在布鲁内尔大学无线通信专业获得工学硕士学位。现在中国科学院大学信息安全专业攻读博士学位, 研究领域为物联网、信息安全, 研究兴趣包括物联网安全、工业控制系统安全。Email: zhangweidong@iie.ac.cn

文辉 于 2016 年在中国科学院大学信息工程研究所信息安全专业获得博士学位。现任中国科学院大学信息工程研究所助理研究员。研究领域为物联网安全、信息安全、网络安全。研究兴趣包括: 恶意代码分析、大数据分析、网络信息探测等。Email: wenhui@iie.ac.cn

朱红松 于 2009 年在中国科学院大学计算所获得博士学位。现任中国科学院信息工程研究所研究员。主要研究方向包括物联网安全、网络攻防、安全大数据分析与测评。Email: zhuhongsong@iie.ac.cn

尹丽波 现任国家工业信息安全发展中心主任。研究领域为网络与信息安全理论于技术研究。研究兴趣包括: 信息安全标准、计算机内容安全、信息安全战略。

孙利民 于 1998 年在国防科技大学计算机体系结构专业获得工学博士学位。现中国科学院信息工程研究所第四研究室研究员。研究领域为物联网安全、工业控制系统安全。研究兴趣包括: 工控入侵诱捕、工控态势感知。Email: sunlimin@iie.ac.cn