IP定位相关工作---朱金玉

一、基于推测的定位算法：一般通过查询whois数据库等来过去该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。

典型算法有IP2LL, NetGe, GeoTrack, VisualRoute, Neotrace, GTrace, Quova, MaxMind, EdgeScape, TraceWare, SoftWare77, IPligence, HostIP, IPInfoDB等。

可分为三类1.直接查询Whois数据库来推测主机位置信息。2.通过测量主机名并结合数据库信息来推测主机地理位置3通过网络结构和数据库信息来推断主机地理位置

文章1：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>

提出NetGeo算法 ，netgeo是映射IP地址，域名，AS到地理位置的工具

Netgeo使用whois得到3中标识符域名，AS号，IP地址找到位置。

文章2：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]

Venkata提出Geotrack算法，认为在主机名字中可能包含了不同粒度的地理信息；Geoping算法，假设到同一主机时延相近的主机，其地理位置也相近；Geocluster是通过数据结构和数据库信息来推断主机地理位置。从而建立一个为网络主机服务的IP地址与地理位置映射。

方法：

1. Geotrack基于DNS域名试着找出目标主机的位置，或其他相近的网络节点。（依赖于最后一只位置的路由与地标之间的距离以及利用路由名定位的准确度）
2. Geoping使用从地理分布位置到相应对等点的目标主机的网络时延测量。（依赖于地标的数量和位置，需要每一个地标都接近每一个目标节点才能保证低的错误位置）
3. Geocluster用BGP前缀信息结合部分IP到位置映射信息推测主机位置。（需要大量的细粒度和新鲜的数据库）

文章3：Periakaruppan R, Nemeth E. Gtrac A Graphical Traceroute Tool. Usenix LISA, 1999.

Gtrace是一个图形化的前端，使用一些启发式的发放来确定节点的位置，Traceroute提供描述地理上的IP路径信息，根据他们的经纬度坐标描述地图上的节点，获得位置。（neotrace visualroute 同）

1. gtrace 执行traceroute到目的主机，通过路径节点决定位置
2. Gtrace验证是否在前一阶段过得的位置合理

二、基于时延的定位算法：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, Shortest Ping，Constraint-based Geolocation（CBG), Topology-based Geolocation (TBG), NBIGA ,Posit, Geo-RX, Spotter 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.

2.1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这5 种算法的基本思想和原理.

文章1：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]

基于约束的定位CBG，使用多点定位地理约束建立连续的空间，推测网络主机的地理位置。将时延精确的转换为距离约束。

1. 多点定位的地理距离约束（准确测量目标点与参考点的距离，利用多个基准点到目标节点的距离采用类似于三角定位的方法确定目标节点所在区域）
2. 从延时测量到距离约束，建立一个动态的关系在网络时延和地理距离间。如果地标之间的延迟在周期性的聚集，会产生一个自校准的算法，每一个地标观察当前网络延迟和网络内的地理距离之间的动态关系。
3. 采用分布式距离约束定位主机
4. 影响的过度和低估的距离约束

文章2：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]

Shortest ping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。

提出基于拓扑的定位TBG，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，如时延的强约束和链路时延弱约束，然后确定待定位节点和中间节点所在的位置区间。

优点：充分利用了距离基准节点越近则定位越准确的，利用中间节点地理位置来检测到目标节点路径的正确性，允许通过迭代达到全局的一致性。

2.2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, Posit, Spotter, 等.

文章1：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]

文章2：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]

1. 综合定位 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, TTG, GeoWeight 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.

文章1;Wong B, Stoyanov I, Sirer EG. Octant: A comprehensive framework for the geolocalization of Internet hosts. In: Proc. of the NSDI2007 Symp. Cambridge: ACM Press, 2007. 313−326. <http://www.cs.cornell.edu/People/egs/papers/octant-nsdi.pdf>

提出Octant一个综合的广泛的框架，使用whois数据库，路由器DNS名称，及无人居住区的已知位置数据可以自然地集成到解决方案中并提炼他。

目标

1. 基于贝塞尔曲线Bezier，提供一个通用基础适合任何地理约束。
2. 有力的约束了网络延时数据到高精准度的位置推测
3. 用planetlab测量主机以及公共路由追踪服务器。精确到22英里

方法

1.映射延迟到距离。计算每个地标到测量节点的时延d,（一个节点i的ping时间为d),距离为R，而r表示排队时延表示的距离。则，R与r表示一个积极的时延与消极的时延，通过三角定位的方法可以得到一个重叠区域。

2.最后一跳的延迟。最后一跳的延迟通常根据配置可以归因于电缆和DSL的连接。a.定位执行迅速，为包膨胀使用精确时间硬件，软件。b创造底层屋里的详细地图网络，c.Octant有以适应不确定性的限制，能负担得起不精确性在最后一跳的延迟估计。

3.间接路由器。 Octant执行一个反转的DNS寻找每一个路由器上的路径，决定它所在的城市通过使用undns工具（从ISP特定命名模式中通过路由器的名字提取位置。）从路由器的城市信息，转换成地理坐标，通过使用US census zipcode 数据库

4.处理不确定性。Octant减少以指数方式增长的延迟，因此当低延迟地标出现时减轻了高延迟地标。加权在起始地标到目标点相关的每一个基于延迟的约束

5.迭代优化。1使用准确和最保守的约束建立一个估计位置区间，包含目标的概率很高；2使用不太精确和更积极的约束获得更好估计区域内部的目标位置初始估计。

6地域限制。除了提取时延测量约束，Octant使任何地理约束被表示成任意的贝塞尔区域，成为定位过程的整体。

1. 点的选择。Octant去定位计算一个最终估计定位区域，捕获目标必须存在的最好的估计。Octant使用Monte-Carlo算法，找一个在最好估计目标区域中的单个点，系统最初随机选择上千个随机点，每个点分配一个权值，在一系列的试验下选择最好的约束地点。

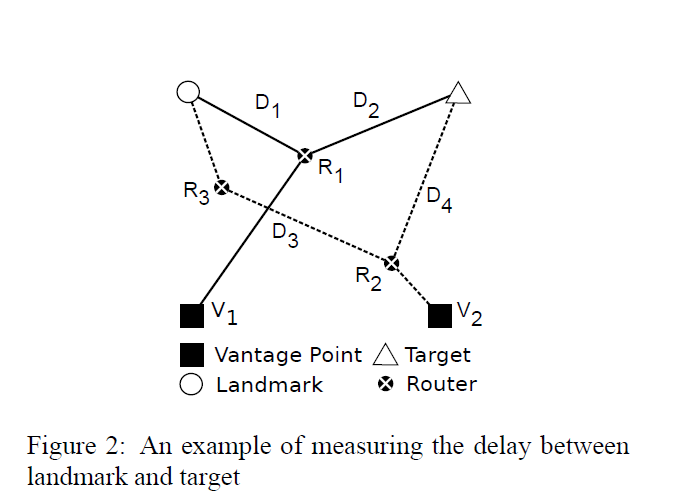
总：Octant框架综合使用各种信息来对目标主机进行定位，他将所有的信息分为积极、消极两类约束条件，以缩小预测区域从而提高定位精度。积极约束条件，通过主动测量来获取被测主机可能在的区域；消极约束条件，计算机不可能的条件，不将可能定位的位置表示成坐标而表示为一个贝塞尔曲线所决定的曲面。在实验不是理想模型时，引入权值表示接入的时延。

文章2;Wang Y, Burgener D, Flores M, Kuzmanovic A, Huang C. Towards street-level client-independent IP geolocation. In: Proc. of the8th USENIX Conf. on Networked Systems Design and Implementation. Berkeley, 2011. https://www.usenix.org/legacy/event/nsdi11/tech/full\_papers/Wang\_Yong.pdf

1我们利用一个系统有效的获取WEB上的定位信息，在一个特定的ZIP code去建立一个地表的数据库

2我们采用了一个三层的系统，从一个大的粗粒度的规模开始，逐步精细、街道水平、规模。它利用地标数据和事实，在一个小规模，相对距离被时延测量保留，克服了使用绝对测量中遇到的许多基本不准确。通过结合这些我们证明了使用积极的延迟测量和网络挖掘的地理位置的目的地的有效性。

1. 确定一个目标IP在一个粗粒度的区域。我们使用一个良好的基于约束的定位方法，稍作修改。发送一个ping服务到目标，将每个ping服务和目标的时延转换为地理距离。我们已知，包在真空，光纤电缆中的速度为光速的2/3，由于传输排队延迟，速度为2/3为上限。我们采用4/9倍的光速作为时延系数。以此我们建立了测量点的距离，以每个测量节点做圆，得到一些覆盖的区域。这些让我们能够找到一个覆盖目标IP的区域。CBG实现误差在143-228km之间。我们查找额外的网络服务信息从ZIPcode中，提取验证位置信息。为了找到ZIPcode自己在属于指定区域中，首先我们在交叉区域中心，以5km做圆，之后采样10个经纬度对在园的周长上，对每个点旋转36度，我们验证他们属于交叉区域
2. 我们利用大量基于网站的基准节点将IP定诶到一个较小范围内，使用traceroute间接估计地标与目标之间的延迟。我们使用越多的traceroute服务器，我们发现地标与目标之间的路径越多。我们首先从所有traceroute服务器发送traceroute探针到地标和目标，为每个vantage point，我们发现最邻近地标和目标的公共路由R1/R2。之后我们计算公共路由器和地标之间的时延D1/D3，和公共路由器到目标的时延D2/D4,我们最后选择地标与目标之间的两个时延的和。从V1点来说时延为D1+D2,从V2来说为D3+D4。



就不同traceroute服务器有不同的路径到目的地，公共路由器不必要与所有的traceroute服务器相同。每一个vantage point 都有一个不同的时延，我们选择最短的时延。再利用ZIPcode。

3 在第二层约束区域上，我们重复采样，我们扩展区域到半径1km，发现所有zipcode 在交叉区域。增加基于网站基准点的数目来选择最近基准点，作为目标位置。

使用不同的数据库 planetlab , residential, online map .