

基于网络坐标的地理信息系统^{*}

王国栋, 陈 阳, 张增斌, 邓北星, 李 星

(清华大学电子工程系下一代网络技术与应用实验室, 北京 100084)

摘 要: 确定网络节点位置是互联网中一种重要的应用服务。目前主流的确定网络节点位置的系统都不易获取且不具有实时性, 给出的节点位置不够准确。提出 1 种基于网络坐标的节点地理位置预测系统, 利用少数地标节点的位置信息, 并结合相应的网络坐标, 对网络当中任意节点的地理位置进行实时地预测。通过在 Internet 网络当中开展实验, 进一步验证了该系统的可行性和准确性。结果表明, 在实验开展的范围内, 系统能够准确的预测节点的地理位置。

关键词: 网络坐标; 地理位置; 经度误差; 纬度误差

中图法分类号: P208

文献标识码: A

文章编号: 1672-5174(2008)05-089-04

近年来, 互联网中产生了很多提供主机地理位置查询的应用, 如腾讯 QQ 中显示用户所在地服务、ip2location 等。而运用主机所在地理位置的信息来优化覆盖网络的方法也在很多网络系统中被应用, 这种做法大大提高了网络的传输效率, 降低了网络的负载。因此, 如何利用网络信息确定 1 台连入 Internet 的主机所在的地理位置有重要的意义。在当今主流的系统, 是通过查询 IP 地址到实际地理位置的映射数据库来确定网络主机的地理位置。虽然这种方法接口简单, 但是数据库的维护存在非常严重的问题。由于 IP 地址分配非常混乱, 想要准确地得到 IP 地址和地理位置的对应关系来建立数据库是比较困难的。由于目前 IP 地址数量有限, 同一地址段往往需要多次重新分配。在 IP 地址重新分配以后, 如何维护数据库也是 1 个难以解决的问题, 数据库更新的不及时将导致难以实时地给出准确的查询结果。

网络坐标技术通过少量的网络测量, 可以准确地预测网络距离, 是 1 种具有高可扩展性的互联网定位方法。网络坐标在 P2P 覆盖网络构建当中有着重要的应用。Vivaldi^[1] 算法是代表性的分布式网络坐标计算方法, 在互联网 P2P 文件共享^[2]、分布式哈希表^[3] 和应用层路由^[4] 中有着广泛的应用。

Michael 和 Karthik 进行的主机网络距离和主机地理距离的相关性研究^[5] 发现, 对大多数 Internet 链路来说, 二者基本成正比例关系。为了实现主机定位的实时性和准确性, 当主机地理位置发生改变的时候及时给出准确的位置信息, 本文提出了 1 种基于网络坐标的主机地理位置预测系统。首先, 人为选定网络当中的少数节点作为地标节点, 获知其地理位置信息(即

经度和纬度)。采用 Vivaldi^[1] 算法实时地计算出地标节点和待定位主机的网络坐标, 利用以上信息通过优化算法确定待定位主机的经度和纬度。网络坐标的计算和节点定位的计算实时进行, 当节点的实际地理位置发生改变时, 预测位置也会相应发生改变。

作者在分布式实验床 Planet Lab^[6] 部署系统并进行实际测量, 采用节点位置预测的经度误差和纬度误差来评价系统的可行性和准确性。

1 基于网络坐标的主机地理位置预测系统设计

系统将节点分成 2 组, 1 组是预先选定的地标节点, 这些节点的地理位置已知, 性能稳定, 不会频繁离开和接入 Internet。另外一组是待定位节点组, 即除地标节点以外的节点。这些节点利用地标节点做参考, 结合本机和地标节点的网络坐标预测自己的地理位置。

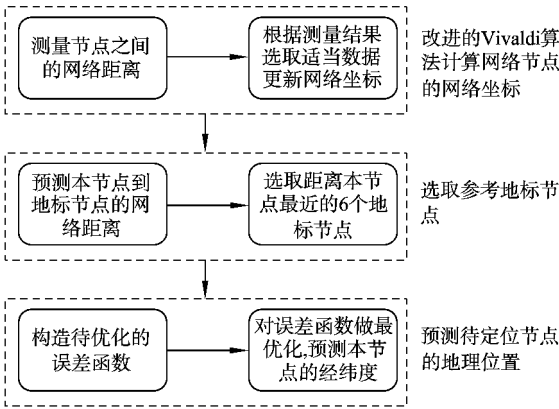


图 1 系统工作的示意图

Fig. 1 The network coordinate based geography information system architecture

* 基金项目: 国家重点基础研究发展规划项目(2007CB310806); 国家自然科学基金(60473087, 60703052)项目资助

收稿日期: 2008-08-20; 修订日期: 2008-09-10

作者简介: 王国栋(1985-), 男, 硕士生。E-mail: wanggd04@gmail.com

系统完成定位的主要过程分为:

- (1) 计算节点网络坐标;
- (2) 选取参考地标节点并且获取地标位置信息;
- (3) 预测主机的地理位置。

系统工作流程示意图见图 1。

1.1 节点网络坐标的计算

系统通过 Vivaldi^[1] 算法计算地标节点和待定位节点的网络坐标。准确的预测节点之间的网络距离是准确预测地理位置的基础。为了尽可能减小网络坐标预测网络距离的误差对于最终地理位置预测误差的影响,文中采用基于滑动窗口机制的 Vivaldi 网络坐标算法,这样可以提高网络坐标计算的准确性^[6]。在这种机制中,需要确定窗口长度 W , 以及窗口内测量值的选取策略。

1.2 选取地标节点以及获取地标节点的位置

完成网络坐标的计算以后,根据节点网络坐标推算出待定位节点到各个地标节点之间的网络距离。然后对这些网络距离排序,选取距离待定位节点最近的 6 个地标节点作为地理位置定位的参考点。下面将利用参考节点的地理坐标进行预测。

1.3 预测主机的地理位置

文中采用 Nelder-Mead 单纯形调优化算法^[7]来预测节点的地理位置。该优化算法需要利用待定位节点到 6 个参考点的网络距离以及 6 个参考点的经纬度为输入,构造 1 个以待定位节点经纬度为自变量的误差函数。

公式(1)所示函数为根据两点经纬度计算地理距离。

$$gDis_{ij} = GeoDistance(Lat_i, Lng_i, Lat_j, Lng_j) \quad (1)$$

优化算法中使用的误差函数如公式(2)。

$$Error(Lat_x, Lng_x) = \sum_{i=1}^6 \rho \left[\frac{gDis_i}{R} - nDis_i \right] \quad (2)$$

误差函数的自变量是待定位节点的经纬度, R 是反映节点之间地理距离和网络距离的比例关系的参数, Michael 和 Karthik 的研究^[1]表明, Internet 当中绝大部分的节点之间的网络距离和他们的地理距离是成正相关,并且这种正相关关系基本对应着正比例的关系,所以 R 代表了地理距离和网络距离之间的正比例关系的比例系数。 $gDis_i$ 和 $nDis_i$ 分别表示的是待定位节点到第 i 个参考点的地理距离和网络距离, ρ 是误差加权系数。

公式(2)中误差的意义是根据给定的节点经纬度,以 6 个参考点为基准,用地理距离预测的网络距离和实际网络距离的误差。待定位节点和参考点之间的网络距离越近,这个参考点对应的误差权重也越大,这样就能让网络上距离待定位节点较近的参考节点在定位

的时候发挥主要的作用。对误差函数运用 Nelder-Mead 单纯形调优化算法^[7]进行优化,找到使得误差函数值最小时的作为输入的最优经度和纬度。这样得到待定位节点的经纬度就是最终预测得到的节点的地理位置。

2 基于网络坐标的节点地理位置预测准确度分析

为了在真实网络环境当中测试本系统的性能,我们在 PlanetLab 实验床上部署了本系统,并开展实验分析节点位置预测的准确度。本实验所选取的 141 节点都分布在北美大陆,地标节点也在这些节点当中选出,这些节点均分布在 $70^\circ W \sim 120^\circ W$, $25^\circ N \sim 45^\circ N$ 之间,实验区域经度跨度 $50(^{\circ})$, 纬度跨度 $20(^{\circ})$ 。

2.1 实验参数的选取

改进的 Vivaldi 算法中采用 8 维网络坐标,滑动窗口长度 $W=5$,在计算网络坐标时选取的是窗口内的 RTT 最小值。本文的实验表明,在这组参数下,网络坐标预测网络距离的误差最小。

在 141 个节点当中选取了 30 个节点作为地标节点,这些节点都分布在北美的高校当中,并且每个高校有且只有 1 个地标节点。

定位时,从 30 个地标节点中选取定位参考地标节点数目是 6。通过反复的实验发现,参考节点选取过多,网络坐标的不准确,会引入较多的误差;参考节点选取太少,会因为个别参考点网络坐标误差过大导致位置预测不准确。

经过实验,确定了实际地理距离和网络距离的比例值 $R=27$,误差加权系数 ρ 和网络距离的平方成反比例关系,即

$$\rho = \frac{1}{nDis_i^2} \quad (3)$$

2.2 节点地理位置预测准确度分析

在这部分中,通过实验获得节点地理位置预测的经度误差和纬度误差,从而反映系统预测地理位置的准确性。

节点位置预测的经度误差和纬度误差定义分别如式(4)和(5)。

$$\text{经度误差} = |\text{节点实际经度} - \text{节点预测经度}| \quad (4)$$

$$\text{纬度误差} = |\text{节点实际纬度} - \text{节点预测纬度}| \quad (5)$$

为了消除因为经度和纬度跨度不同造成的平均误差大小不同,进一步定义相对平均误差,如式(6)和(7)所示。

$$\text{相对平均经度误差} = \frac{\text{平均经度误差}}{\text{经度跨度范围}} \times 100\% \quad (6)$$

$$\text{相对平均纬度误差} = \frac{\text{平均纬度误差}}{\text{纬度跨度范围}} \times 100\% \quad (7)$$

图 2 和 3 分别是实验统计得到的经度误差和纬度误差的累计分布曲线。

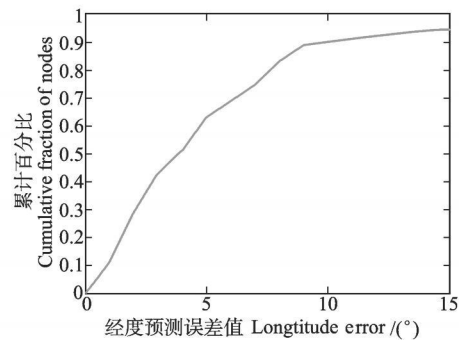


图 2 经度预测误差累计分布曲线图

Fig.2 The cumulative distribution of 141 plantlab node's longitude error

从图 2 可以看到, 大约有 64% 的节点的经度预测误差是在 $5(^{\circ})$ 以内。实验得到 141 个节点经度预测误差的平均值是 $5.067\ 212(^{\circ})$ 。由于实验节点经度跨度是 $50(^{\circ})$, 因此代入式(6) 得到经度预测的相对平均误差约为 10.1%。

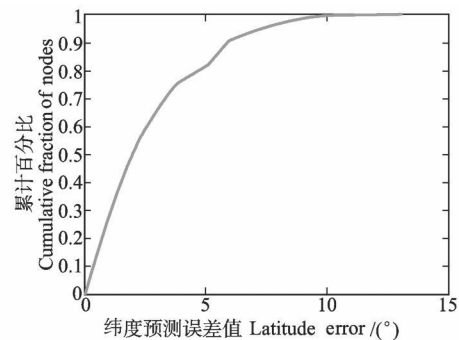


图 3 纬度预测误差累计分布曲线图

Fig.3 The cumulative distribution of 141 planetlab node's latitude error

从图 3 可以看到, 大约有 80% 的节点的纬度预测误差是在 $5(^{\circ})$ 以内的。对于 141 个节点纬度预测误差的平均值是 $2.779\ 825\ 3(^{\circ})$ 。由于实验节点纬度跨度是 $20(^{\circ})$, 因此代入式(7) 可以得到纬度预测的相对平均误差约为 13.9%。

尽管受到网络坐标的误差, 网络拥塞变化的随机性以及网络路由机制的复杂性影响, 但是不论是经度预测还是纬度预测, 相对平均误差都在 14% 以内, 实验表明节点定位结果是比较准确的。图 4 是在电子地图上实现的基于网络坐标的地理信息演示系统界面。图中的圆形点表示的是系统定位得到的主机位置, 方框表示考虑到定位误差, 主机实际位置可能存在的地理范围。

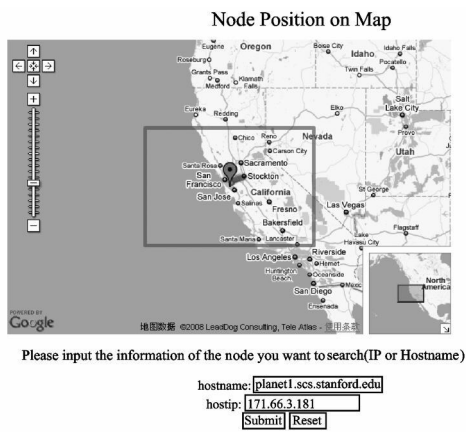


图 4 基于网络坐标的地理信息演示系统界面

Fig.4 A demo of network coordinate based geography information system

4 结语

本文针对 Internet 中节点位置查询的实时性要求, 提出了 1 种基于网络坐标的网络节点地理位置预测方法, 并在此方法基础上构建了基于网络坐标的地理信息系统。进一步将该系统部署到 PlanetLab 上开展大量基于实际互联网的实验, 实验表明, 在开展实验的范围内, 该系统预测的节点地理位置是比较准确的。我们设计实现的系统为节点地理位置查询服务提供了新的解决方案。

参考文献:

[1] Frank Dabek, Russ Cox, Frans Kaashoek, et al. Vivaldi: A decentralized network coordinate system [C]. USA: ACM SIGCOMM' 04 Conference, Portland, Oregon, 2004.

[2] Ledlie J, Gardner P, Seltzer M. Network coordinates in the wild [C]. USA: NSDI' 07, Cambridge, MA, 2007.

[3] Sean Rhea, Dennis Geels, Timothy Roscoe, et al. Handling churn in a DHT [C]. USENIX Annual Technical Conference, Boston: Massachusetts, 2004.

[4] Ledlie J, Pietzuch P, Mitzenmacher M, et al. Wired geometric routing [C]. Washington USA: IPTPS, Bellevue, 2007.

[5] Michael J. Freedman, Karthik Lakshminarayanan, et al. OASIS: Anycast for any service [C]. San Jose, CA, USA: NSDI' 06, 2006.

[6] Peter Pietzuch, Jonathan Ledlie, Margo Seltzer. Supporting network coordinates on planetLab [C]. San Francisco, CA, U SA: WORLDS 2005.

[7] Nelder J A, Mead R. A simplex method for function minimization [J]. Computer Journal, 1965, 7: 308-313.

Network Coordinate Based Geographic Information System

WANG Guo-Dong, CHEN Yang, ZHANG Zeng-Bin, DENG Bei-Xing, LI Xing

(Lab of New Generation Network T&A, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

Abstract: Locations of the hosts are very important for many Internet applications. However, this location information is hard to obtain in existing service, either not real time or lack of accuracy. In this paper, a Network Coordinate based host locating system is presented, in which, hosts' geographic positions are derived from their network coordinates and geographic information of the landmarks. Our extensive experiments on PlanetLab show that our system can predict hosts' geographic positions accurately.

Key words: network coordinates; geographic position; longitude error; latitude error

责任编辑 徐 环

(上接 84 页)

Analysis of Development and Trends of the Utilization of IP Phone

HU Ping, HUANG Yong-Feng

(Network Research Center, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: This paper describes the global development and the trends discussions of utilization of IP Phone. And the current solutions by Chinese government and IP Phone service providers in China are introduced. This analysis will supply references to Chinese government in making strategic plan for future utilization of IP Phone in China.

Key words: IP phone; internet; status; analysis

责任编辑 庞 旻