



(12)发明专利

(10)授权公告号 CN 105282267 B

(45)授权公告日 2018.08.14

(21)申请号 201510585592.2

(22)申请日 2015.09.15

(65)同一申请的已公布的文献号
申请公布号 CN 105282267 A

(43)申请公布日 2016.01.27

(66)本国优先权数据
201510545711.1 2015.08.31 CN

(73)专利权人 刘粉林
地址 450001 河南省郑州市高新区科学大
道62号信息工程大学
专利权人 罗向阳

(72)发明人 刘粉林 朱光 张刚 陈晶宁
赵帆 罗向阳

(74)专利代理机构 郑州联科专利事务所(普通
合伙) 41104

代理人 刘建芳 李伊宁

(51)Int.Cl.
H04L 29/12(2006.01)

(56)对比文件
CN 1985186 A,2007.06.20,
CN 101031144 A,2007.09.05,
CN 101548197 A,2009.09.30,
CN 102405420 A,2012.04.04,
US 5745682 A,1998.04.28,
CN 104715012 A,2015.06.17,
EP 1062825 A1,2000.12.27,

审查员 牟茜茜

权利要求书2页 说明书4页 附图2页

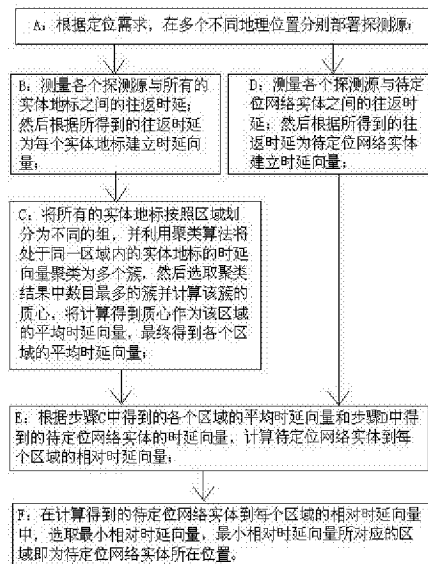
(54)发明名称

一种基于地标聚类的网络实体城市级定位方法

(57)摘要

本发明公开了一种基于地标聚类的网络实体城市级定位方法,包括以下步骤:A:在多个不同地理位置分别部署探测源;B:测量各个探测源与所有的实体地标之间的往返时延;C:将所有的实体地标按照区域划分为不同的组,得到各个区域的平均时延向量;D:测量各个探测源与待定位网络实体之间的往返时延;E:计算待定位网络实体到每个区域的相对时延向量;F:选取最小相对时延向量,最小相对时延向量所对应的区域即为待定位网络实体所在位置。本发明与基于网络测量的网络实体定位技术相比,能够有效避免网络拥塞、负载均衡、异构网络和网络设备性能较差等因素对网络时延膨胀和时延抖动造成影响,显著提高实体地标挖掘的数量,同时提高网络实体城市级的定位精度,可为网络城市级定位提供可靠地标。

CN 105282267 B



1. 一种基于地标聚类的网络实体城市级定位方法,其特征在于:包括以下步骤:

A: 根据定位需求,在多个不同地理位置分别部署探测源;然后分别进入步骤B和步骤D;

B: 测量各个探测源与所有的实体地标之间的往返时延;然后根据所得到的往返时延为每个实体地标建立时延向量;然后进入步骤C;

C: 将所有的实体地标按照区域划分为不同的组,并利用聚类算法将处于同一区域内的实体地标的时延向量聚类为多个簇,然后选取聚类结果中数目最多的簇并计算该簇的质心,将计算得到质心作为该区域的平均时延向量,最终得到各个区域的平均时延向量;然后进入步骤E;

D: 测量各个探测源与待定位网络实体之间的往返时延;然后根据所得到的往返时延为待定位网络实体建立时延向量;然后进入步骤E;

E: 根据步骤C中得到的各个区域的平均时延向量和步骤D中得到的待定位网络实体的时延向量,计算待定位网络实体到每个区域的相对时延向量;然后进入步骤F;

F: 在计算得到的待定位网络实体到每个区域的相对时延向量中,选取最小相对时延向量,最小相对时延向量所对应的区域即为待定位网络实体所在位置。

2. 根据权利要求1所述的基于地标聚类的网络实体城市级定位方法,其特征在于:所述的步骤B中,利用不同地理位置部署的探测源分别向网络中所有的实体地标发送ICMP探测数据包,测量各个探测源与所有的实体地标之间的往返时延,将所有的探测源与每一个网络实体地标之间的往返时延记为 $d_{i1}, d_{i2}, d_{i3}, \dots, d_{im}$, $i=1, 2, \dots, m$,其中 i 表示第 i 个探测源,并分别为每个地标建立时延向量 DV , $DV = (d_{i1}, d_{i2}, d_{i3}, L, d_{im})$, $i=1, 2, \dots, m$ 。

3. 根据权利要求2所述的基于地标聚类的网络实体城市级定位方法,其特征在于:所述的步骤C中,将所有的实体地标以城市为单位划分为不同的组,采用K-means算法将处于同一区域内的实体地标的时延向量聚类为多个簇,然后选取聚类结果中数目最多的簇并计算该簇的质心,将计算得到质心作为该区域的平均时延向量 $ADV = (delay_1, delay_2, \dots, delay_m)$,其中,K-means算法中的 K 为经验门限值, $delay_1, delay_2, \dots, delay_m$ 为 m 个探测源分别测量得到的时延向量,最终得到各个区域的平均时延向量。

4. 根据权利要求3所述的基于地标聚类的网络实体城市级定位方法,其特征在于:所述的步骤D中,利用不同地理位置部署的探测源分别向待定位网络实体发送ICMP探测数据包,测量各个探测源与待定位网络实体之间的往返时延;然后根据所得到的往返时延为待定位网络实体建立时延向量 DV_target ,

$DV_target = (d'_{i1}, d'_{i2}, \dots, d'_{im})$, $i=1, 2, \dots, m$;其中 i 表示第 i 个探测源。

5. 根据权利要求4所述的基于地标聚类的网络实体城市级定位方法,其特征在于:所述的步骤E中,根据步骤C中得到的各个区域的平均时延向量和步骤D中得到的待定位网络实体的时延向量,采用相对时延计算方法分别计算待定位网络实体到每个区域的相对时延向量 RDV ,

相对时延计算公式如下:

$RDV = (|delay_1 - d'_{i1}|, |delay_2 - d'_{i2}|, \dots, |delay_m - d'_{im}|)$, $i=1, 2, \dots, m, i \leq m$,其中 i 表示第 i 个探测源。

6. 根据权利要求5所述的基于地标聚类的网络实体城市级定位方法,其特征在于:所述的步骤F中,从计算得到的待定位网络实体到每个区域的相对时延向量中选取最小相对时

延向量,最小相对时延向量所对应的城市即为待定位网络实体所在城市,即待定位网络实体的最终城市级定位结果。

一种基于地标聚类的网络实体城市级定位方法

技术领域

[0001] 本发明涉及信息安全技术领域,尤其涉及一种基于地标聚类的网络实体城市级定位方法。

背景技术

[0002] 网络实体定位,也称为IP定位,指确定IP地址对应的网络实体在地理空间中的所在位置,其实质是获取网络实体IP地址与此网络实体地理位置之间的映射关系。由于掌握IP地址的地理位置可以为用户提供各种具有针对性和个性化的高效服务,因此,近年来IP定位技术逐渐受到商业公司、政府机关甚至是个人的关注,例如:商业公司可根据IP地址的地理位置有针对性的向目标群体推送广告和提供基于位置的服务(Location Based Service,LBS)等服务;政府机关可以通过IP地址的地理位置对不同区域推送本地天气预报和自然灾害预警等信息;个人也可根据IP地址的地理位置判断信用卡诈骗行为、处理垃圾邮件及改善P2P(Peer to Peer)网络性能等。

[0003] 基于网络测量的网络实体定位技术是获取高精度定位结果的主要方式之一。然而,网络拥塞、负载均衡、异构网络和网络设备性能较差等因素往往会造成网络时延膨胀和时延抖动等问题,这些因素都会影响基于时延测量的定位方法精度。因此,如何减少上述情况对定位结果的影响,以及如何较为准确的描述某一区域的平均时延是提高基于测量的网络实体定位精度的关键之一。

发明内容

[0004] 本发明的目的是提供一种基于地标聚类的网络实体城市级定位方法,能够利用聚类算法求出的区域内平均时延,以实现高精度的网络实体定位。

[0005] 本发明采用下述技术方案:

[0006] 一种基于地标聚类的网络实体城市级定位方法,其特征在于:依次包括以下步骤:

[0007] A:根据定位需求,在多个不同地理位置分别部署探测源;然后分别进入步骤B和步骤D;

[0008] B:测量各个探测源与所有的实体地标之间的往返时延;然后根据所得到的往返时延为每个实体地标建立时延向量;然后进入步骤C;

[0009] C:将所有的实体地标按照区域划分为不同的组,并利用聚类算法将处于同一区域内的实体地标的时延向量聚类为多个簇,然后选取聚类结果中数目最多的簇并计算该簇的质心,将计算得到质心作为该区域的平均时延向量,最终得到各个区域的平均时延向量;然后进入步骤E;

[0010] D:测量各个探测源与待定位网络实体之间的往返时延;然后根据所得到的往返时延为待定位网络实体建立时延向量;然后进入步骤E;

[0011] E:根据步骤C中得到的各个区域的平均时延向量和步骤D中得到的待定位网络实体的时延向量,计算待定位网络实体到每个区域的相对时延向量;然后进入步骤F;

[0012] F:在计算得到的待定位网络实体到每个区域的相对时延向量中,选取最小相对时延向量,最小相对时延向量所对应的区域即为待定位网络实体所在位置。

[0013] 所述的步骤B中,利用不同地理位置部署的探测源分别向网络中所有的实体地标发送ICMP探测数据包,测量各个探测源与所有的实体地标之间的往返时延,将所有的探测源与某一个网络实体地标之间的往返时延记为 $d_{i1}, d_{i2}, d_{i3}, \dots, d_{im}, i=1, 2, \dots, m$,其中 i 表示第 i 个探测源,并分别为每个地标建立时延向量 $DV, DV=(d_{i1}, d_{i2}, d_{i3}, L, d_{im}), i=1, 2, \dots, m$ 。

[0014] 所述的步骤C中,将所有的实体地标以城市为单位划分为不同的组,采用K-means算法将处于同一区域内的实体地标的时延向量聚类为多个簇,然后选取聚类结果中数目最多的簇并计算该簇的质心,将计算得到质心作为该区域的平均时延向量 $ADV=(delay_1, delay_2, \dots, delay_m)$,其中,K-means算法中的K为经验门限值,最终得到各个区域的平均时延向量。

[0015] 所述的步骤D中,利用不同地理位置部署的探测源分别向待定位网络实体发送ICMP探测数据包,测量各个探测源与待定位网络实体之间的往返时延;然后根据所得到的往返时延为待定位网络实体建立时延向量 DV_target ,

[0016] $DV_target=(d'_{i1}, d'_{i2}, \dots, d'_{im}), i=1, 2, \dots, m$;其中 i 表示第 i 个探测源。

[0017] 所述的步骤E中,根据步骤C中得到的各个区域的平均时延向量和步骤D中得到的待定位网络实体的时延向量,采用相对时延计算方法分别计算待定位网络实体到每个区域的相对时延向量 RDV ,

[0018] 相对时延计算公式如下:

[0019] $RDV=(|delay_1-d'_{i1}|, |delay_2-d'_{i2}|, \dots, |delay_m-d'_{im}|), i=1, 2, \dots, m, i \leq m$,其中 i 表示第 i 个探测源。

[0020] 所述的步骤F中,从计算得到的待定位网络实体到每个区域的相对时延向量中选取最小相对时延向量,最小相对时延向量所对应的城市即为待定位网络实体所在城市,即待定位网络实体的最终城市级定位结果。

[0021] 本发明首先利用不同地理位置的探测源获得各个探测源与所有的实体地标之间的往返时延RTT,然后根据所得到的往返时延RTT为每个实体地标建立时延向量 DV ,再将实体地标按区域划分后利用聚类算法获得各个区域的平均时延向量 ADV ,极大地提高了各个区域平均时延的精度。随后,本发明利用不同地理位置的探测源获得各个探测源与待定位网络实体之间的往返时延RTT,并建立待定位网络实体时延向量 DV_target ,再根据各个区域的平均时延向量 ADV 和待定位网络实体的时延向量 DV_target 分别计算得出待定位网络实体到每个区域的相对时延向量 RDV ,最后,本发明从计算得到的待定位网络实体到每个区域的相对时延向量 RDV 中选取最小相对时延向量 RDV ,最小相对时延向量 RDV 所对应的区域即为待定位网络实体所在位置。本发明与基于网络测量的网络实体定位技术相比,能够有效避免网络拥塞、负载均衡、异构网络和网络设备性能较差等因素对网络时延膨胀和时延抖动造成影响,显著提高实体地标挖掘的数量,同时提高网络实体城市级的定位精度,可为网络城市级定位提供可靠地标。

附图说明

[0022] 图1为本发明的流程示意图;

[0023] 图2为相对时延计算方法的原理示意图。

具体实施方式

[0024] 以下结合附图和实施例对本发明作以详细的描述：

[0025] 如图1所示，本发明所述的基于地标聚类的网络实体城市级定位方法，主要包括聚类部分和定位部分。聚类部分包括步骤A至步骤C；定位部分包括步骤D至步骤F。本发明选取城市级基准节点，选取同一ISP (Internet Service Provider) 下的多个区域城市级网络实体地标作为定位算法的基准节点。

[0026] 本发明所述的基于地标聚类的网络实体城市级定位方法，依次包括以下步骤：

[0027] A: 根据定位需求，在多个不同地理位置分别部署探测源；本实施例中，可以根据定位精度需求，在不同地理位置部署m个探测源。

[0028] B: 利用不同地理位置部署的探测源分别向网络中所有的实体地标发送探测报文，测量各个探测源与所有的实体地标之间的往返时延RTT (Round-Trip Time)；然后根据所得到的往返时延RTT为每个实体地标建立时延向量DV (Delay Vector)。

[0029] 步骤B中，可利用不同地理位置部署的探测源分别向网络中所有的实体地标发送ICMP探测数据包，测量各个探测源与所有的实体地标之间的往返时延RTT，将所有的探测源与某一个实体地标之间的往返时延记RTT为 $d_{i1}, d_{i2}, d_{i3}, \dots, d_{im}, i=1, 2, \dots, m$ ，其中i表示第i个探测源，并分别为每个实体地标建立时延向量DV， $DV = (d_{i1}, d_{i2}, d_{i3}, \dots, d_{im}), i=1, 2, \dots, m$ ；

[0030] C: 将所有的实体地标按照区域划分为不同的组，并利用聚类算法将处于同一区域内的实体地标的时延向量DV聚类为多个簇，然后选取聚类结果中数目最多的簇并计算该簇的质心，将计算得到质心作为该区域的平均时延向量ADV (Average Delay Vector)，得到各个区域的平均时延向量ADV。

[0031] 本发明中，将所有的实体地标以城市为单位划分为不同的组，采用K-means算法将处于同一城市内的实体地标的时延向量DV聚类为多个簇，然后选取聚类结果中数目最多的簇并计算该簇的质心，将计算得到质心作为该区域的平均时延向量 $ADV = (delay_1, delay_2, \dots, delay_m)$ ；

[0032] $ADV = \text{Centroid}(\text{MaxCluster}(\text{cluster}_1, \dots, \text{cluster}_k))$

[0033] $= (delay_1, delay_2, \dots, delay_m)$ ；

[0034] 其中，Centroid()为质心计算函数，MaxCluster()为最大簇获取函数， $\text{cluster}_1, \dots, \text{cluster}_k$ 分别为聚类产生的K个簇， $delay_1, delay_2, \dots, delay_m$ 为m个探测源分别测量得到的时延向量。

[0035] K-means算法为成熟的现有算法，在此不再赘述。K-means算法中的K是一个经验门限值，能够通过大量实验并结合用户对定位精度的需求而确定。聚类算法及质心的计算同样属于现有技术。

[0036] D: 利用不同地理位置部署的探测源分别向待定位网络实体发送ICMP探测数据包，测量各个探测源与待定位网络实体之间的往返时延RTT；然后根据所得到的往返时延RTT为待定位网络实体建立时延向量DV_target；

[0037] $DV_target = \text{GeTargetRTT}(\text{probes}, \text{target}) = (d'_{i1}, d'_{i2}, \dots, d'_{im}), i=1, 2, \dots, m$ ；

[0038] 其中，GeTargetRTT()为得到目标的往返时延函数，probes为探测源，target为待

定位网络实体, $d'_{i1}, d'_{i2}, \dots, d'_{im}$ 分别为 m 个探测源测量到的与待定位网络实体之间的往返时延, i 表示第 i 个探测源。

[0039] E: 根据步骤C中得到的各个区域的平均时延向量ADV和步骤D中得到的待定位网络实体的时延向量DV_target, 计算待定位网络实体到每个区域的相对时延向量RDV (relative Delay Vector);

[0040] 本发明中, 采用相对时延计算方法分别计算待定位网络实体到每个区域的相对时延向量RDV,

[0041] 相对时延计算公式如下:

[0042] $RDV = (|delay_1 - d'_{i1}|, |delay_2 - d'_{i2}|, \dots, |delay_m - d'_{im}|), i = 1, 2, \dots, m, i \leq m$, 其中 i 表示第 i 个探测源;

[0043] F: 在计算得到的待定位网络实体到每个区域的相对时延向量中, 选取最小相对时延向量, 最小相对时延向量所对应的区域即为待定位网络实体所在所在城市, 即待定位网络实体的最终城市级定位结果。

[0044] 最小相对时延向量的选取方法属于现有技术, 可通过计算待定位网络实体到每个区域的相对时延向量的平均数, 并依据平均数大小进行选取。

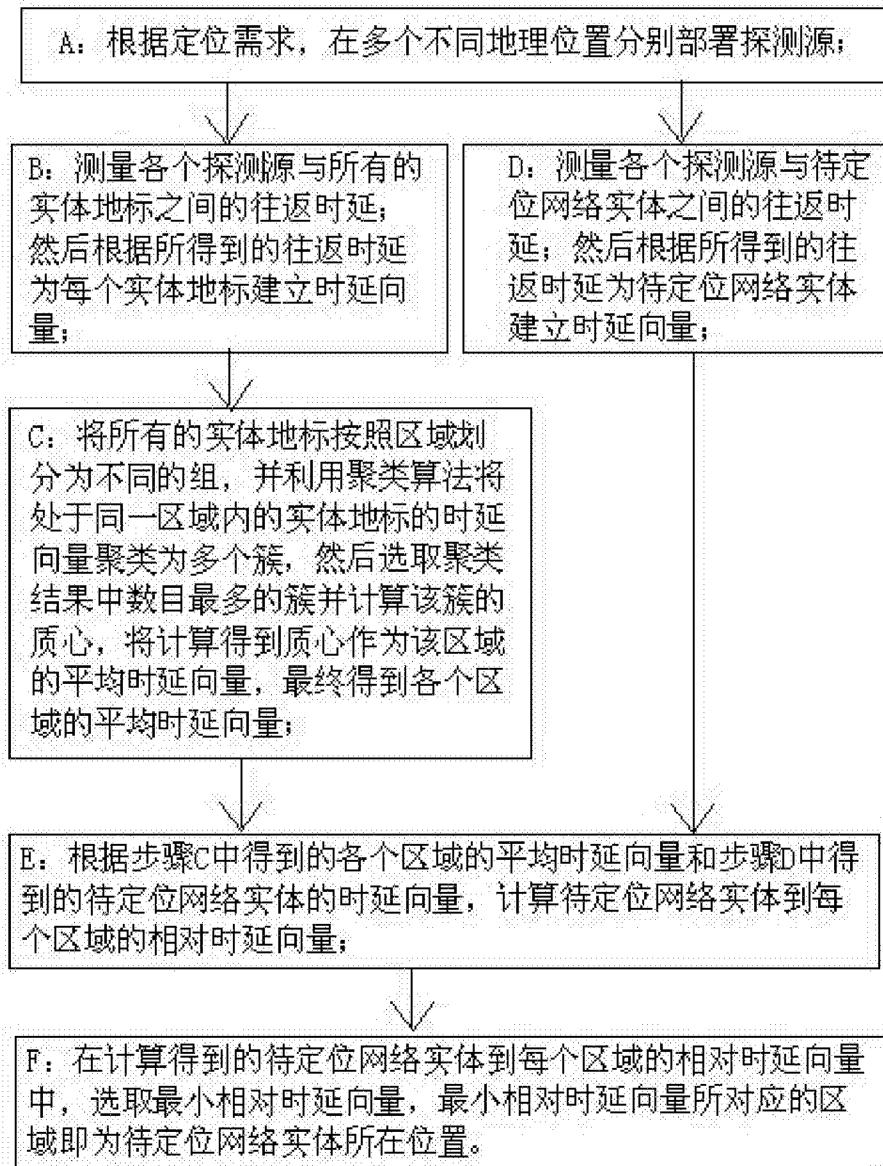


图1

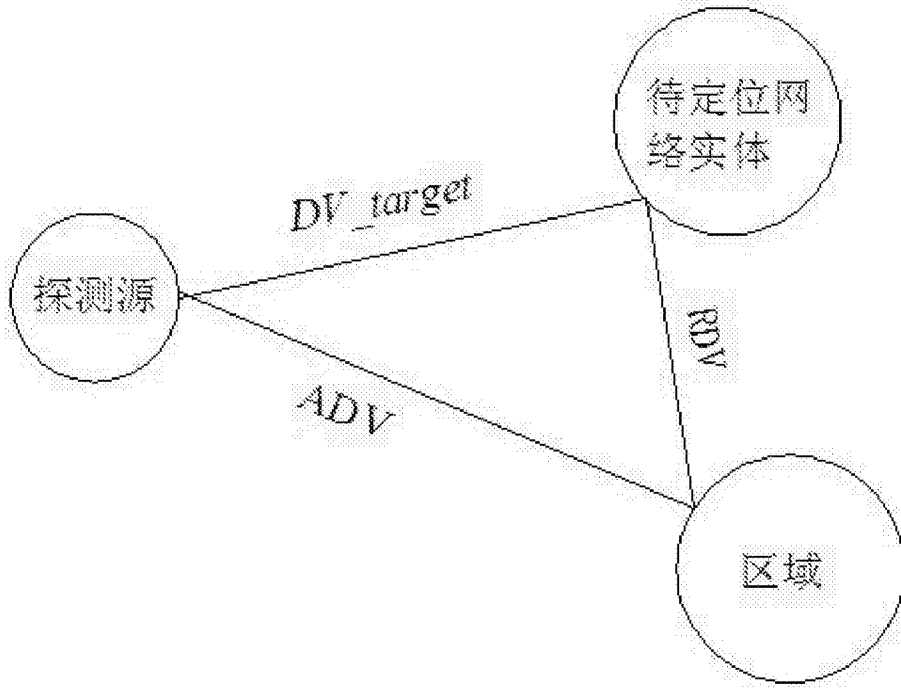


图2