# 一种P2P网络中基于位置感知的节点选择策略

刘永贤　王洪波　程时端　林　宇
北京邮电大学网络与交换技术国家重点实验室　北京　100876

摘　要　P2P网络作为一种覆盖网络，邻居节点的选择若不考虑网络层、物理层信息，将导致较低的数据传输速度和不必要的跨运营商流量，从而大大限制P2P技术的应用。位置感知策略可以解决这些问题。本文对现有位置感知策略进行分类，并通过分析比较指出了现有位置感知策略的优缺点。提出了一种基于IP地址库的简单有效的节点选择策略，能够感知节点的运营商信息和物理位置。

关键词　P2P网络；节点选择；位置感知；IP地址库

## 引言

目前，随着互联网的发展和普及，P2P（Peer to Peer）技术已经成为被广泛关注的技术。P2P网络称为对等网络，它打破了传统的Client/Server模式，在对等网络中，每个节点的地位都是相同的，具备客户端和服务器双重特性，可以同时作为服务使用者和服务提供者。P2P技术在文件下载、流媒体、即时通信等方面得到了广泛的应用。

P2P网络是一种覆盖网络(Overlay Network)，覆盖网络是指建立在另一个网络上的网络，简单说来就是在现有的因特网上构建一个完全位于应用层的网络系统，它是面向应用层的，不考虑或很少考虑网络层、物理层的信息。

P2P网络中一个重要的概念是邻居节点，即与当前节点有应用层连接的节点，在拓扑上是互相连接的。如果当前节点和某个邻居节点有数据交换，则此二者之间连接的质量就会变得非常重要，因为这将直接关系到用户体验。但是另一方面，在当前的覆盖网络中，几乎不能确保这种应用层连接与底层的网络层拓扑的一致性，这将导致以下两方面的问题：

1）较低的数据传输速度。在进行数据交换的时候，节点间的距离往往会成为影响传输速度的一个因素。与远端节点进行数据交换，由于路由路径的加长、各种排队时延的增加等因素均会导致其速度将远小于和近距离节点的数据交换速度，这将直接影响用户的应用体验，特别是P2P流媒体等实时应用。在这种情况下，希望能通过避免这种不必要的远端节点带来的高延迟路由来改进数据传输性能。这样，高性能的大范围P2P网络，需要在构造P2P网络时吸收网络层拓扑信息，来改进数据传输性能。

2）不必要的跨运营商间流量。当前P2P网络的实现忽略了运营商(ISP)链接的代价，从而使得P2P系统大大增加了跨运营商的流量，增加了运营商的经营成本。这将造成一个恶性循环：一方面，网络流量特别是骨干网流量的大幅增加迫使运营商采取限制的手段，强制减少用户的P2P流量，但是这将会导致用户的不满；另一方面，由于运营商之间端口的带宽有限，跨运营商的流量必然也会造成节点间的数据传输延时较大，影响用户的使用效果，从而用户希望建立更多的连接以提高数据传输速度，而这又将招致运营商的不满。由此可见，减

资助课题：中兴通讯研究基金、高等学校博士学科点专项科研基金(No.200800131019)、新世纪优秀人才支持计划(No.NECT-07-0109)

少跨运营商的P2P流量是非常必要的，在构建P2P网络时需要考虑运营商信息。

在P2P网络中加入位置感知策略可以有效解决以上问题。在P2P网络中引入位置感知，将达到P2P业务提供商、用户、网络运营商三方多赢的效果：一方面可以提高P2P应用的服务质量，提高用户体验，另一方面也就提高了P2P业务提供商的用户量，为其实现盈利打下基础，同时还可以减少运营商间的流量，降低运营商运营成本。因此，关于P2P网络位置感知策略的研究意义重大，并已成为学术和产业界当前的研究和关注热点。

本文首先对覆盖网络中现有的各种位置感知策略进行分析，然后提出了一种简单易实现的基于位置感知的邻居节点选择策略，能够感知节点的物理位置和运营商的信息。第二节对各种已有的位置感知策略的理论进行分析，并进行了评价；第三节提出了一种基于IP地址库的简单可行的邻居节点选择策略；最后一节对全文进行总结。

## 1 位置感知策略研究

### 1.1 概述

互联网节点的位置感知并不是一个刚出现的问题，在覆盖网络风靡互联网之前，针对Web服务器的服务器选择也是此问题的一个重要的应用领域。本节研究了目前位置感知问题的几种主要的解决方案，包括运营商参与的流量本地化、基于时延测量的位置感知策略、基于网络拓扑的位置感知策略和基于地理位置的位置感知策略等。

### 1.2 运营商参与的流量本地化

位置感知的目的是实现流量本地化，也就是节点间的数据交换尽量在本自治域(Autonomous System，AS)内，尽量减少跨自治域的数据流量和跨运营商网络的数据流量。

实现这一目的的最好方法就是网络运营商的参与[1]，网络运营商在自己的网络边缘部署一些代理服务器，这些服务器可以将一些数据请求转发到本自治域内的一些具有相应资源的节点，从而实现将流量控制在本自治域。如图1，灰色节点表示A在自治域内的邻居节点，1表示当客户端A的邻居节点无法满足服务要求时向代理服务器请求数据，代理服务器会优先在自治域内查找节点，2表示代理服务器的返回消息，包含了自治域内的可用节点B，3表示A向客户端B请求数据。近年来更是兴起了一种P4P[2]的理论，能够实现更加复杂的网络运营商和P2P应用程序之间的交互和协商，从而实现二者利益的最大化。这种方式固然是最好的方式，但是需要运营商的协助，并且部署代价和维护代价都很大。

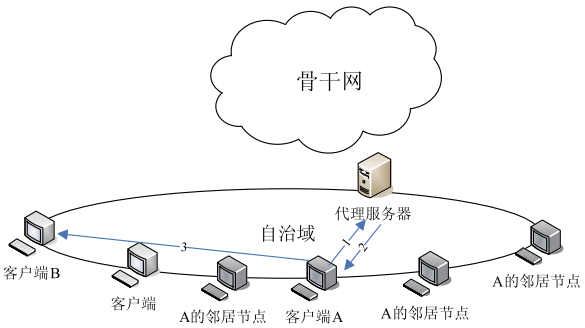



图1 在网络边缘部署代理服务器使流量尽量在本地

### 1.3 基于时延测量的位置感知策略

目前的位置感知理论中，较为常见的就是基于时延测量的位置感知策略，其基本原理就是测量两个节点之间的往返时延(round-trip time，RTT)，将其作为二者网络距离的唯一度量。

#### 1.3.1 网络坐标

在基于时延测量的位置感知策略中，网络坐标是最为常见的实现方式。在节点规模较大时，为了判断出两个节点之间的网络距离而进行两两通信，势必会造成较大的网络通信代价，自然的想法是不必每次都让两个节点直接通信。网络坐标理论就是为了解决这一问题，它的基本原理是建立一个虚拟坐标系，为网络中每个节点分配一个虚拟坐标来表示它们的位置，两个节点之间无须直接通信，而是通过二者的坐标值推算出它们之间的往返时延。

1) 基于标杆节点的网络坐标

这种网络坐标的特点是，依据系统中已有的一些节点作为标杆节点(Landmark)，事先计算好标杆节点的坐标值，后加入的节点通过计算到这些标杆节点的

RTT来计算出自己的坐标值。这种方法又分几种：

① 相对坐标。直接以节点到多个标杆节点的RTT作为坐标，按照一定的算法来计算节点间的距离。这样的策略如triangulated heuristic[3].

② 绝对坐标，集中式标杆。将网络中所有的节点映射到一个绝对的几何空间，每个节点分配一个绝对的几何坐标，通过计算节点间的几何距离来推算节点间的网络距离。集中式标杆是指，整个网络中部署几个专门的标杆节点以供其他节点参考，所有普通节点只能通过到这些集中的标杆节点的距离来计算自己的坐标。在这种策略中，系统首先计算出标杆节点间的相互距离，进而为它们分配初始的坐标值，如图2，然后当新节点加入之后，计算此新节点到所有标杆节点的距离，然后按照一定的最小化算法如单形体下坡(Simplex Downhill)计算出此新节点的坐标值，如图3。在这样的策略中，集中部署的标杆节点势必会成为系统的瓶颈，使得系统的可扩展性变差。这样的策略如GNP[4](Global Network Positioning).

③ 绝对坐标，分布式标杆。同集中式标杆一样，节点映射到绝对的几何空间，所不同的是，这里的标杆节点并不是固定的几个，空间中所有的节点均可以作为标杆，只要此节点的坐标已经确定。这样的策略比前一种灵活，如PIC[5](Practical Internet Coordinates).

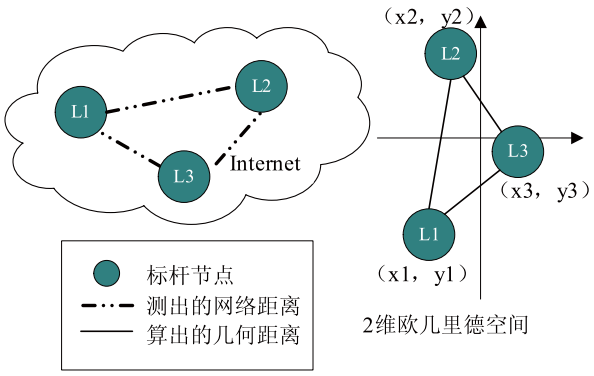



图2 标杆节点计算自己的坐标

2) 基于模拟技术的网络坐标

这一类系统将整个网络中的节点看做一个物理系统，最有名的就是Vivaldi[6]。这种理论将网络看做一个弹簧系统，所有邻居节点之间的都认为具有一个弹簧，将两个节点之间坐标系距离和RTT之间的误差值看做弹簧的伸缩量，如果这个值不为零，则根据物理学中的虎克定理，弹簧会对两端施加一个与伸缩量有关的力，节点会在这个力的作用下移动一小段距离，从而使得这个误差值变小，也就是使得坐标更加准确。Vivaldi实际的算法要复杂得多，考虑了很多因素，如每个节点都有多个邻居，所以节点是在多个力的合力下移动；采用附加式(piggy-back)模式来测算节点间的RTT值，不再造成额外的网络负载；为自身坐标误差调整设定一个本地误差因子，限制坐标调整的幅度；因为邻居节点的坐标也不一定是准确的，所以为邻居节点的坐标设定一个远端误差因子。

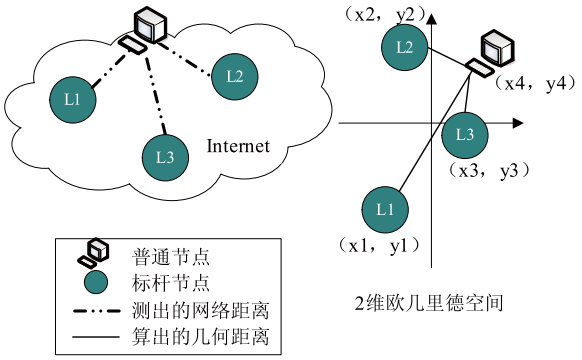



图3 新加入节点根据标杆节点计算自己的坐标

Vivaldi还讨论了坐标空间对于网络坐标系统的影响，提出一种叫做高度向量(Height vectors)的坐标空间，并指出这是一种精确度比较高的坐标空间。这种空间把骨干网部分看作是平面的欧几里德空间，而把接入网部分看作是建立在欧几里德平面上的高度向量。和两端是接入网中间是骨干网这样的网络模型一样，两个节点之间的网络距离等于两端的高度向量和中间的欧几里德距离之和。

和其他位置感知理论不同的是，Vivaldi目前有

较大规模的应用。例如，在开源的BitTorrent客户端Azureus中，就使用了Vivaldi作为它的位置感知优化的机制。

Vivaldi也有其自身的缺点：由于新加入节点的初始坐标是任意给的，比如原点，所以它的坐标收敛的速度不够快，对于大量新加入节点的处理效果不会很好。

#### 1.3.2 分簇

这种策略[7]是将互联网上所有的节点分成一些簇(bin)，认为簇内节点间的距离较近，而这里所说的距离指的是网络延迟，也即RTT。这种策略也需要网络中存在一些节点作为标杆节点，新加入节点测得和这些标杆节点的距离然后决定加入哪个簇，如图4所示。

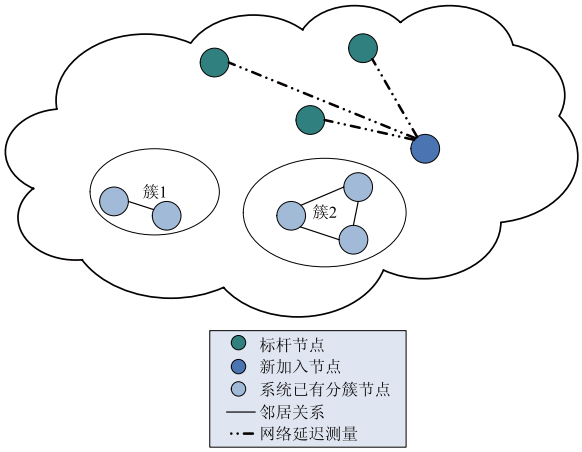



图4 节点根据到标杆节点的距离进行分簇

#### 1.3.3 IDMaps

IDMaps[8]也采用一些标杆节点(这里叫做tracer)，但是这些标杆节点的作用不是建立网络坐标，而是靠他们之间的网络距离来估算普通节点间的网络距离。其基本方法是：每个标杆节点测量到其它标杆节点间的RTT，然后测量它们到每个CIDR地址前缀的RTT，并对每个这样的地址前缀选出离它最近的的标杆节点。对于两个普通主机h1和h2，它们之间的RTT可以被估算为以下三者之和：h1的前缀到此前缀的标杆节点的RTT，h2的前缀到此前缀的标杆节点的RTT，然后是这两个标杆节点之间的RTT。

#### 1.3.4 小结

本节分析的是基于时延测量的位置感知策略，这是一种非常常见的方法。因为相比其他因素网络延迟是网络距离的直接度量，所以这些方法都具有一定的意义。但是，基于时延测量的方法也有它的缺点。首先，准确性较低，由于网络拥塞等一些原因，这些方法的准确性都难以达到很高的水平；另外，此类方法把网络看做一个黑盒子，无视其拓扑结构，可能会导致一些问题，比如对网络拓扑的改变反应较差。

### 1.4 基于网络拓扑的位置感知策略

要想把数据流量限制在自治域内，掌握网络的拓扑结构是一个很重要的方法，但是由于网络运营商都不愿意将自己网络的详细拓扑公布出来，只能通过一些技术手段来获取网络拓扑信息。以下是拓扑感知方法。

#### 1.4.1 基于网络前缀匹配的拓扑感知

由于网络前缀(network prefix)和子网的划分是有关系的，所以从主机的网络前缀可以大致推算出主机的子网关系，这样可以通过网络前缀匹配来进行节点的位置感知。这种方法还可以分为简单匹配和分层匹配，简单匹配是指选用固定的位数作为网络前缀的位数来进行匹配选择，而分层匹配是分别采用几个不同的位数来进行匹配，比如先用24位，然后用16位、14位等。

这种方法最大的问题就是准确性，由于自治域常常不使用连续的地址空间，使得基于前缀匹配方法的有效性大大降低。

#### 1.4.2 基于路由表信息的拓扑感知

这种方法[9]通过分析位于核心网路由器的路由表来建立网络的拓扑信息，其依据是路由表项里的网络前缀和子网掩码能够标识出路由路径，通过将这些表项进行分簇大致可以归纳出自治域的分布。虽然研究表明[1]，这种方法可以达到和2.2所述方法接近的准确度，然而，这种方法需要获取大量路由器的路由表，操作代价太高，并且一旦网络拓扑发生变化更新拓扑图的代价也很高。

#### 1.4.3 基于traceroute的拓扑感知

由于traceroute可以返回所经路径上所有设备的IP地址，所以利用它来获知网络拓扑也成为一种很重要的手段，目前有很多这样的工具，如skitter，RIPENCC TTM等，其基本原理是从预设的一些源节点发送消息到很多目标节点，通过返回的路径设备信息来获知网络拓扑。这种方法也有很多弊端。首先，如果要对整个网络的拓扑有较为清晰的了解，需要较多的源节点。其次，如果有大规模的消息指向同一个目标节点，则很容易造成分布式拒绝服务攻击(Distributed denial of Service，DDoS)。还有，这样大规模的消息本身会造成网络负载加重。最后，在网络拓扑发生变化的情况下对网络拓扑图的维护会是一个问题。

### 1.5 基于地理位置的位置感知策略

研究表明[10]，节点间的物理距离和网络延迟之间是有很大关系的，互联网的连通性越强时这种现象就越加明显。由此可见，以节点间的物理距离来估算网络距离是合理的，可以看作是轻量化的位置感知策略。

节点的距离然后决定加入哪个簇，如图4所示。

## 2 基于IP地址库的节点选择策略

### 2.1 概述

由第二节的分析，我们可以看出：虽然目前存在很多种节点位置感知的策略，但是绝大多数都存在一定的问题，理论准确度高的实施难度大，而实施难度小的准确率又很难保证，如基于时延的位置感知策略根本无法感知网络的拓扑结构，所以也无法感知网络运营商信息，无法减少跨运营商的网络流量。

本文提出了一种简单易实现的位置感知策略，其基本思路是使用IP地址库。由于IP地址库包含了IP地址段的地域信息和运营商信息，由2.5的分析可知，地域信息可以用来估算网络距离，而运营商信息本身就是一种网络拓扑信息，可以作为参考进行节点选择。IP地址库本身存在不够准确和过期的问题，对于前者，由于物理位置并不能完全代表网络延迟，但是考虑到这种方法实现的简单性，从性价比和可行性的角度考虑是可以接受的；对于后者，只要采用一定的技术手段，这个问题是可以解决的。

### 2.2 运营商标识

由于IP地址库里带有运营商的信息，直接将其提取出来即可使用，在具体实现方面，可以为每个运营商编一个号。如果IP地址库不够完善，可能有的IP地址段无法标识运营商信息，这样可以为这些IP地址段统一编一个号。

### 2.3 节点位置标识

本文所讨论的节点选择策略主要是针对国内的互联网用户，把全国的地域按照城市进行划分，按照IP地址库将各个IP地址段映射到相应的城市。

如何表示城市的地理位置是一个问题。一种办法是为全国建立一个绝对的平面坐标系，然后为每个城市分配一个坐标。这种方法准确度不高，因为地球是球面的，平面坐标系表示法必然有很大误差。本文采用球面坐标系即经度和纬度来标识城市。

通过经纬度计算两点间的大地距离采用地理坐标系两点间的距离公式：

$D = \arccos[\sin\phi 1\sin\phi 2+\cos\phi 1\cos\phi 2\cos(\lambda 2 \lambda 1)]\ R$（$\lambda$表示经度，$\phi$表示纬度，D为大地线距离，R为地球半径）

### 2.4 节点选择策略

#### 2.4.1 节点位置信息向量

本文把节点的经度、纬度和运营商代码三个信息的组合称为“节点位置信息向量”，此信息向量可以通过各种方式被邻居节点所掌握，作为节点选择的依据。

#### 2.4.2 IP映射库

首先要对原始IP地址库进行处理，由于原始IP地址库里表示位置信息的域分别是文字形式的城市和运营商，如图5所示，需要将这两个域分别映射为经纬度和运营商代码，并将原始IP地址库里的无用信息剔除，这样就形成了可以被系统使用的IP映射库。

这样的IP映射库可以放在某个服务器上，如Tracker或者专门的服务器。当新节点加入系统时立即与服务器通信，服务器通过此节点的外部IP地址查询IP

映射库获得此节点的信息向量，然后回传给此新加入节点。位置信息向量可以作为节点信息的一部分保存在本地，而节点信息以后可以通过各种方式被邻居节点所掌握，作为节点选择的依据，整个系统的简要流程图如图6。将IP映射库放在服务器上可以很好的进行控制，很容易进行更新等操作。

### 2.4.3 节点选择

前文的基本假设是节点可以掌握其邻居节点的信息，包括位置信息向量，这在实现上并不困难。当节点需要进行邻居节点选择时，如图6所示，可以按照以下方法：

首先，将此节点的邻居节点按照位置信息向量中的运营商代码进行排序，同运营商的节点排在最前面，然后是不同运营商的邻居节点。这种做法是基于同运营商优先于物理距离这样一个原则，尽量将数据流量限制在同一个运营商的网络中。

然后，对于同运营商网络内的节点，根据其经度和纬度计算其到本节点的距离，按照其到本节点的距离远近进行排序。

BBSGOOD_IP

| ip1 | ip2 | ip3 | ip4 | country | city |
|---|---|---|---|---|---|
| 1032163327 | 1032164350 | 61.133.144. | 61.133.147. | 安徽省蚌埠市 | 网通 |
| 1032164351 | 1032165630 | 61.133.148. | 61.133.152. | 安徽省阜阳市 | 网通 |
| 1032165631 | 1032165886 | 61.133.153. | 61.133.153. | 安徽省淮南市 | 网通 |
| 1032165887 | 1032166142 | 61.133.154. | 61.133.154. | 安徽省马鞍山 | 网通 |
| 1032166143 | 1032166398 | 61.133.155. | 61.133.155. | 安徽省淮南市 | 网通 |
| 1032166399 | 1032167422 | 61.133.156. | 61.133.159. | 安徽省淮北市 | 网通 |
| 1032167423 | 1032168190 | 61.133.160. | 61.133.162. | 安徽省宿州市 | 网通 |
| 1032168191 | 1032168702 | 61.133.163. | 61.133.164. | 安徽省巢湖市 | 网通 |
| 1032168703 | 1032168958 | 61.133.165. | 61.133.165. | 安徽省六安市 | 网通 |
| 1032168959 | 1032169214 | 61.133.166. | 61.133.166. | 安徽省滁州市 | 网通 |
| 1032169215 | 1032169470 | 61.133.167. | 61.133.167. | 安徽省安庆市 | 网通 |
| 1032169471 | 1032169982 | 61.133.168. | 61.133.169. | 安徽省安庆市 | 电信 |
| 1032169983 | 1032170494 | 61.133.170. | 61.133.171. | 安徽省安庆市 | 网通 |
| 1032170495 | 1032172030 | 61.133.172. | 61.133.177. | 安徽省马鞍山 | 网通 |
| 1032172031 | 1032172542 | 61.133.178. | 61.133.179. | 安徽省宣城市 | 网通 |
| 1032172543 | 1032173054 | 61.133.180. | 61.133.181. | 安徽省黄山市 | 电信 |
| 1032173055 | 1032173760 | 61.133.182. | 61.133.184. | 安徽省铜陵市 | 电信 |

图5 IP地址库示例

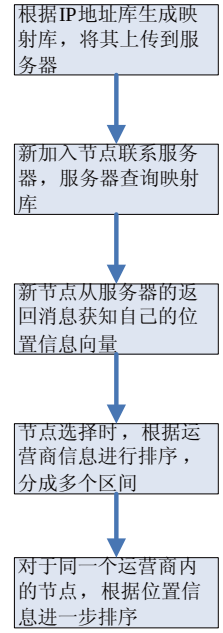



图6 节点选择整体流程图

## 3 总结

P2P网络中对节点物理位置信息的忽视将导致较低的数据传输速度和不必要的跨运营商流量这两方面的问题，将大大限制P2P这种优秀技术的应用，而位置感知策略可以解决这些问题。这些问题的解决将对P2P业务提供商、用户、网络运营商三方产生多赢的效果。

本文首先对现有的位置感知策略进行了总结，在此基础之上提出了一种简单易行的基于位置感知的节点选

择策略，主要是将运营商信息和物理位置作为节点选择的主要考虑因素，以IP地址库作为策略的信息来源，具有较高的实用价值。

本文提出的策略也存在一些不足：IP地址库存在粗粒度的问题，对于IP地址库过期的同步也是一个需要考虑的问题，另外集中式的服务器管理IP映射库存在瓶颈问题，这些问题都可以在以后的研究中继续加以解决和改善。

## 参考文献


[1] Thomas Karagiannis, Pablo Rodriguez, Konstantina Papagiannaki. Should Internet Service Providers Fear Peer-Assisted Content Distribution?. IMC '05, October 2005

[2] Haiyong Xie, Arvind Krishnamurthy, Avi Silberschatz.et al.P4P: Explicit Communications for Cooperative Control Between P2P and Network Providers. P4P Working Group Whitepaper, July 2007

[3] S.M. Hotz. Routing information organization to support scalable interdomain routing with heterogeneous path requirements, Ph.D.Thesis (draft), University of Southern California, 1994

[4] T.S.E.Ng,Zhang.H.Predicting Internet network distance with coordinates-based approaches//Proceedings of IEEE Infocom, 2000:170-179

[5] Costa M.,Castro M.,Rowstron A.,et al. PIC:Practical Internet coordinates for distance estimation//International Conference on Distributed Systems,Tokyo,Japan,March 2004

[6] Dabek F.,Cox R.,Kaashoek F.,et al. Vivaldi:A decentralized network coordinate system.SIGCOMM'04,Aug.30 Sept.3,2004

[7] Shenker S.,Ratnasamy S.,Handley M., et al.Topologically-aware overlay construction and server selection.Proceedings of INFOCOM'02,2002

[8] Francis P.,Jamin S.,Jin C.,et al. IDMaps: A global Internet host distance estimation service.IEEE/ACM Transactions on Networking, October 2001

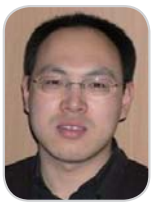
[9] Krishnamurthy B.,Wang J.On network-aware clustering of web clients.In Proceedings of ACM Sigcomm,August 2000

[10] Padmanabhan V.,Subramanian L.An investigation of geographic mapping techniques for Internet hosts//Proceedings of ACM SIGCOMM,San Diego,August 2001:173-185

## 作者简历


**刘永贤**

硕士研究，北京邮电大学网络技术研究院网络与交换技术国家重点实验室生。目前研究方向为P2P技术。



**王洪波**

博士，北京邮电大学网络技术研究院网络与交换技术国家重点实验室，副教授，硕士生导师。目前研究方向为互联网测量与管理、分布式计算、下一代互联网体系结构及新应用等。

**程时端**

博士生导师，北京邮电大学网络技术研究院网络与交换技术国家重点实验室教授。目前研究方向为宽带网络性能和服务质量、流量工程、下一代网体系结构、P2P技术等。

**林　宇**

博士，北京邮电大学网络技术研究院网络与交换技术国家重点实验室，副教授。研究方向为互联网服务质量管理与测量、P2P计算等授，博士生导师。目前研究方向为宽带网络性能和服务质量、流量工程、下一代网体系结构、P2P技术等。


# A Locality-Aware Based Method for Peer Selection in P2P Network


Liu Yongxian
Wang Hongbo
Cheng Shiduan
Lin Yu

State Key Laboratory of Networking and Switching, Beijing University of Posts & Telecommunications, Beijing 100876, China



**Abstract** If neighbors' information involving network layer and locality is scarcely taken into account in Peer Selection in P2P network, two problems of low transmission rate and unnecessary cross-ISP traffic will be caused, which will then lead to a limit to the spread of P2P technology. Locality awareness can solve these problems. This paper classifies related locality-aware methods, pointing out the advantages and disadvantages of each one by comparison and analysis, then proposes an easy-to-implement and effective method for peer selection based on IP address base, which considers neighbors' information of ISP and locality.



**Keywords** P2P Network; Peer Selection; Locality-Aware; IP Address Base