

因特网时延空间建模的研究

王占丰, 陈鸣, 邢长友, 李兵, 仇小锋

(解放军理工大学 指挥自动化学院, 江苏 南京 210007)

摘 要: 首先概述了因特网时延空间建模的基本概念, 然后从因特网时延空间模型演化的角度对典型模型进行了分析, 剖析因特网时延空间建模中存在的难点。最后, 对现有的因特网时延空间模型进行了综合对比, 指出了建模趋势及主要研究方向。

关键词: 网络测量; 因特网时延空间; 网络性能; 网络坐标

中图分类号: TP393

文献标识码: A

文章编号: 1000-436X(2012)07-0164-13

Research on the modeling of the Internet delay space

WANG Zhan-feng, CHEN Ming, XING Chang-you, LI Bing, QIU Xiao-feng

(Institute of Command Automation, PLA University of Science & Technology, Nanjing 210007, China)

Abstract: Firstly, the concept of the modeling of the Internet delay space was presented, and then various delay models were analyzed from the evolvement view. In the following section, the challenges of Internet delay space modeling were pointed out. Finally, the future trends of the modeling of the Internet delay space were discussed.

Key words: network measurement; Internet delay space; network performance; network coordinate

1 引言

因特网已经成为现代社会的基础设施, 网络性能的优劣备受因特网服务提供者(ISP, Internet service provider)和用户的关注。特别是近些年来, 大量时延敏感型网络应用不断涌现, 如 P2P、网格计算、云计算等。在这些应用中, 由于能够提供相同服务的服务器不止一个, 用户在请求某项服务时可以从选择一个具有较小时延的服务器, 从而大大减少通信时间。采用传统的网络测量方式来选择服务器需要进行大量的测量, 不仅增加了网络开销而且还不能提供实时信息。人们希望建立一个因特网时延模型来准确地预测任意 2 个节点间的时延。因此, 因特网时延空间的建模成为网络测量领域最为活

跃的研究方向之一, 取得了众多的理论和实践成果^[1]。一些具有代表性的网络时延模型和系统被构建出来, 并得到广泛部署和应用。如著名的 P2P 客户端软件 Azureus 就使用了坐标系统 Vivaldi 来选择最近的对等节点^[2], 在微软游戏 Xbox LIVE 中也使用了坐标系统 Htrae 来选择较近的服务器^[3]。

因特网时延的研究始于因特网建立之初, 现在已有网络协议支持和众多可供使用的测量工具。最初, 其研究对象以排队时延、链路时延为主, 如今则侧重于整个网络因特网时延空间的分析和建模。在 1994 年, Hotz 在其博士论文中首次对因特网时延空间中的三角形不等式约束进行了研究^[4], 开启了因特网时延空间研究的先河。Ng 等^[5]在 2002 年提出了首个网络坐标系统 GNP, 自此之后因特网时

收稿日期: 2011-08-08; 修回日期: 2012-02-28

基金项目: 国家重点基础研究发展计划“973”计划基金资助项目(2012CB315800); 国家自然科学基金资助项目(61070173); 江苏省自然科学基金资助项目(BK2009058, BK2010133)

Foundation Items: The National Basic Research Program of China (973 Program) (2012CB315800); The National Natural Science Foundation of China (61070173); The Natural Science Foundation of Jiangsu Province (BK 2009058, BK 2010133)

延空间的研究异常活跃,许多学者提出了不同网络坐标系统和因特网时延空间模型^[2~18]。网络坐标系统的提出改变了传统的因特网时延空间认知方式。人们开始从全网络的角度研究因特网时延的分布规律,以虚拟的数学空间来进行建模和指导网络测量实践。在网络坐标系统中,通过测量节点到少数基准节点间的时延,用空间嵌入算法为每个节点分配一个坐标值 $(H_1, H_2, H_3, \dots, H_N)$ 。这样,网络应用需要感知网络时延时,仅需要知道节点自身坐标和目的节点坐标就可以通过距离公式计算出节点之间的时延,从而大大减少网络测量开销,因而备受关注。

本文全面阐述了因特网时延空间建模的基本概念,分析了主要因特网时延空间模型依据和基本思想,指出了因特网时延空间建模的难点,最后对未来研究趋势进行探讨。

2 因特网时延空间建模的基本概念

2.1 因特网时延空间的基本性质

时延是衡量网络性能优劣的主要参数之一,关于因特网时延的研究一致备受关注。现在网络中已经有专门的网络协议(ICMP)和大量的测量工具支持时延测量(如 ping、traceroute、king、Turbo-king^[19]、owping^[20]等)。IETF 下属的 IP 网络性能工作组^[21]在 RFC 2330 中对时延的测量方法及采样等制定了详细的规范。近些年来,一些大的网络测量平台(如 PlanetLab、DIMES^[22]、OneLab、EmuLab 等)和测量项目(如 Skitter、ARK、HPS³等)的出现,使得大规模时延数据的获得更为容易,为时延空间的研究奠定了基础。

时延一般分为单向时延和往返时延 2 类,在因特网时延空间建模中使用的是往返时延。

定义 1 因特网时延空间 对于一个具有 n 个节点 $N=\{N_1, N_2, N_3, \dots, N_n\}$ 的网络,任一节点 N_i 到网络中所有节点的时延(往返时延)组成一个 n 维距离向量 $D_i=(D_{i1}, D_{i2}, \dots, D_{in})$,将所有节点的距离向量所构成的空间 $D=\{D_1, D_2, \dots, D_n\}$ 称为网络时延空间 相应地,由因特网节点的距离向量所构成的空间称为因特网时延空间。

因特网时延空间的性质是因特网时延空间建模的基础,也是网络测量研究的一个重要方向。以往的研究涉及到了时延的各方面性质,包括时延的非对性、抖动性、随时间的变化规律等,但是在时延空间建模过程中主要关注的有 3 点性质,即因特

网时延空间是否为线性空间、度量空间特性以及空间维数特征(增长规律)。通过长期测量,其性质归纳如下。

1) 在网络节点地理距离较远时,节点间的时延与地理距离成线性关系^[3],而节点间距离较小时,则 2 者线性关系不明显^[14,23]。

2) 因特网时延空间中存在着大量的违反三角形不等式(TIV, triangle inequality violation)现象(即网络中任意 3 个节点,其中 2 个节点间的时延大于它们各自另外 1 个节点时延的和),使得因特网时延空间不能成为一个度量空间^[24,25]。

3) 因特网时延空间具有分形特征,其维数不能用整数维表示,因而因特网时延空间不能用同一的空间模型表示^[26]。

因特网时延空间的上述特征说明因特网时延空间是不均匀的,各个部分并不具有同一性质。这种不均匀性就为因特网时延空间的模型选择、嵌入维数确定带来了困难,特别是 TIV 现象与人们的常识不符,也使得许多以度量空间模型为基础的坐标系存在机理性的问题。

2.2 因特网时延空间建模的基本概念

定义 2 等距同构 对任意 2 个测度空间 $M_1(X, \delta)$ 和 $M_2(Y, \mu)$ 而言,若存在一个映射 $f: X \rightarrow Y$, 使得

$$\exists r > 0, \forall x, y \in X (r\delta(x, y) \leq \mu(f(x), f(y)) \leq Sr\delta(x, y)) \quad (1)$$

则称空间 M_1 能够以扭曲度 S 嵌入到空间 M_2 中。其中, S 越接近于 1, 则称该嵌入扭曲越小。特别地,若 $S=1$, 那么称该嵌入为等距同构(isometry)。

定义 3 因特网时延空间建模是指为因特网时延空间建立准确的数学模型来描述因特网时延空间的拓扑结构。在数学意义上,就是寻找与因特网时延空间 D 的等距同构空间。其思想是构造一个网络节点 N 到 M 维实向量的映射 $f: N \rightarrow R^M$, 将网络节点 $N=\{N_1, N_2, N_3, \dots, N_n\}$ 映射为 M 维几何空间中的 n 个坐标点 $H=\{H_1, H_2, H_3, \dots, H_N\}$, 同时使得根据坐标值计算得出的距离与实际测量得到的时延误差值最小,如式(2)所示。

$$D_{ij} \approx \hat{D}_{ij} = \|H_i - H_j\|, \forall i, j = 1, 2, \dots, N \quad (2)$$

例如,在欧氏空间中,节点的距离可以表示为

$$\hat{D}_{ij} = \|H_i - H_j\| = \left(\sum_{k=1}^M (H_{ik} - H_{jk})^2 \right)^{1/2}, \forall i, j = 1, 2, \dots, N \quad (3)$$

其中, \hat{D}_{ij} 代表节点 H_i 和 H_j 之间的时延, H_i 表示空间坐标。图 1 给出了将网络节点嵌入到一个 3 维欧氏空间的示意。

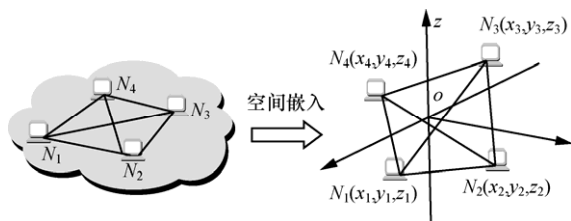


图 1 基于坐标系统的因特网延迟空间模型

因特网延迟空间建模的过程大致可以分为 3 步: 1) 根据网络因特网延迟空间的性质, 选择一个尽可能与其等距同构的几何空间; 2) 分析因特网延迟空间的固有维数, 以确定嵌入空间的维数; 3) 选择适当的嵌入算法, 为每个节点分配空间坐标。

2.3 因特网延迟空间模型的分类

因特网延迟空间的模型可以按照不同方法进行分类。如按照嵌入空间的类型划分, 可以分为基于度量空间的模型和基于非度量空间的模型; 按照嵌入算法的设计原理分类, 可以分为基于模拟的嵌入算法、基于数学计算的算法等; 按照基准节点的选择可以分为随机选择和非随机选择。为了更好地理解因特网延迟空间建模的研究规律, 本文按照因特网延迟空间模型提出和演化过程进行研究, 将其分为 3 类。

1) 单空间模型(single model): 即将因特网延迟空间 D 映射到一个目标空间 R^M 的模型, 为网络中每个节点分配一个唯一的坐标, $f: D \rightarrow R^M$ 。

2) 复合空间模型(compound model): 在复合模型中, 因特网延迟空间 D 被划分为多个子空间和一个核心子空间 $D=D_1 \cup D_2 \dots D_n \cup D_c$, 并且将不同的子空间 X_i 映射到不同的目标空间 Y_i 中, 因此 D 中的每个元素拥有 2 个坐标, 即一个局部坐标和一个全局坐标, $f_i: D_i \rightarrow R^M_i$, 当预测位于不同子空间的 2 个节点间时延时, 通过核心空间来进行叠加 $g(h_1, h_2) = g(h_1, h_1^c) + g(h_1^c, h_2^c) + g(h_2, h_2^c)$ 。

3) 层次空间模型(hierarchical model): 在多空间模型中, 因特网延迟空间 D 被划分为多个子空间 $D=D_1 \cup D_2 \dots D_n$, 并且把不同的子空间 X_i 映射到不同的目标空间 R^M_i 中, 因此 D 中的每个元素拥有多个坐标, $f_i: D_i \rightarrow R^M_i$ 。

单空间模型出现在因特网延迟空间建模的初

始阶段, 不同的学者从不同角度提出了基于各种原理, 适用于不同数据集的嵌入算法, 主要解决了嵌入算法的设计问题。在后 2 个阶段, 则着重提高嵌入精度。这一过程反映了人们对因特网延迟空间性质认识的日益深入, 模型更加精细, 体现了网络的连接结构。由于模型很多, 表 1 列出了在较为重要的期刊和会议提出的主要模型以及分类。

2.4 因特网延迟空间建模与网络距离预测的区别

因特网延迟空间建模与网络距离预测是一对十分容易混淆的概念。因特网延迟空间建模着重描述时延的分布规律和性质, 反映网络在不同时刻的性能状况, 以及节点间的相互关系。它可以被用于预测节点间的时延, 但并不是唯一的用途, 因特网延迟空间的模型还可以用于恶意节点的检测、寻找最短路径等。如文献[27]利用节点间的时延作为指纹特征来确定主机的相对位置, 以检测主机哄骗。Cristian 深入研究了 TIV 的成因和特点, 并利用 TIV 来寻找最优路径[28]。

网络距离预测是利用节点自身某种或几种属性作为测度来衡量节点间的距离[1]。虽然时延是最常用的一种距离测度, 但却不是唯一的测度, 如节点 IP 地址、节点间的可用带宽以及分组丢失率也常常被用来作为网络距离测度。IDMaps[29]中基于 IP 地址前缀来判断网络节点的距离, 认为具有相同地址前缀的主机在地理上也比较接近。BRoute[30]、Sequoia[31]、SABI[32]、PathGuru[33]把可用带宽作为节点选择服务提供节点的依据, 而 Queen[34]则是通过分组丢失率来度量节点间的距离。

本文的研究范围仅限于因特网延迟空间建模的内容, 包括了网络距离预测中与延迟空间建模的相关部分, 而不包括网络距离预测的其他内容。一些结构化的时延预测方法和网络邻近度估计系统不在本文的研究范围之内, 如 iPlane[36]及其改进 iPlane Nano、Meridian、Navigator 等。

3 单空间模型

单空间模型的前提是因特网延迟空间各个部分具有同一的性质。由于网络节点间的时延与地理位置有关, 因此人们假设因特网延迟空间同一个欧氏空间等距同构, 与度量空间十分接近, 且各个部分差异较小。典型的单空间模型有 GNP、Virtual Landmarks[7]、ICS(Internet coordinate system)[8]、PIC[9]、Vivaldi[13]、LAT[15]、IDES[11]、DS[211]、BBS-E[37]

表 1 不同模型的对比

模型名称	选择策略	测量开销	类型	嵌入空间	部署方式	安全性	TIV 避免
GNP	随机	$O(L^2+LH)$	单	欧氏	集中	No	No
ICS	非随机	$O(L^2+LH)$	单	欧氏	集中	No	No
Lighthouse	非随机	$O(L^2+LH)$	混合	欧氏	分布	No	No
PIC	非随机	$O(L^2+LH)$	单	欧氏	集中	Yes	No
BBS-H	随机	$O(L^2+LH)$	单	双曲	集中	No	No
BBS-E	随机	$O(L^2+LH)$	单	欧氏	集中	No	No
Vivaldi	随机	—	混合	欧氏	分布	Yes	No
IDES	非随机	$O(L^2+LH)$	单	—	集中	Yes	No
HA	随机	$O(VN+LH)$	多	欧氏	集中	Yes	No
Pharos	随机	$O(N)$	多	欧氏	集中	No	No
NetPharos	随机	$O(VN+LH)$	混合	欧氏	集中	No	No
LAT	随机	$O(L^2+LH)$	混合	欧氏	分布	Yes	Yes
Sequoia	—	$O(N^2)$	单	Treeness	集中	No	Yes
HE	随机	$O(VN+LH)$	多	双曲+ 欧氏	集中	No	Yes
DS ²	随机	$O(N)$	单	欧氏	集中	Yes	Yes
TIV aware	随机	$O(L^2+LH)$	单	欧氏	分布	Yes	Yes
Toread	随机	$O(VN+LH)$	多	欧氏	分布	Yes	Yes
Vivaldi-T	随机	$O(N)$	多	欧氏	分布	Yes	Yes
Phoenix	随机	$O(L^2+LH)$	单	—	分布	No	Yes
DMF	随机	$O(L^2+LH)$	单	—	分布	No	Yes
Hierarchical Vivaldi ^[35]	随机	$O(VN+LH)$	多	欧氏	分布	No	No
HNPS, HBBS and HIDES	随机	$O(VN+LH)$	多	欧氏, 双曲	分布	No	No
Htrae	随机	—	单	欧氏	集中	No	No
Phoenix	随机	$O(VN+LH)$	多	—	分布	No	No

L : 基准节点数目 ; H : 普通节点数目 ; N : 节点总数 ; V : 全局基准节点数目

以及 BBS-H^[38]等。单空间模型是因特网时延空间建模的基础，许多复杂的网络坐标系统都是在单空间模型的基础上组合得到的。下面，分析几种典型的单空间模型的嵌入算法，主要包括典型的 GNP 算法、基于模拟的嵌入算法与基于数学计算的嵌入算法。

3.1 GNP

Ng^[5]开创性地提出了通过虚拟坐标嵌入进行网络距离预测的机制 GNP。在该机制中，将时延空间建模为一个欧氏空间，网络中每个节点被映射为欧氏空间中的一个坐标点，按照位置关系为每个节点分配一个坐标值，不同节点间距离可以根据坐标值运算得出。通过上述变换，将网络距离问题转换成了空间节点之间的距离问题。

根据 GNP 的思想，网络节点包括 2 类：基准节点(landmark)和普通节点。其中，基准节点是按

照一定准则在网络中选出的一些代表性节点，其他互相测量到达对方的距离值，然后根据测量结果，最小化实际测量值和按照虚拟坐标计算值之间的误差来确定各个基准节点的坐标值。式(4)描述了优化目标函数表达式，其中， $\varepsilon(D_{ij}, \hat{D}_{ij})$ 表示节点 H_i 和 H_j 之间预测误差评价函数。

$$f_{\text{obj1}}(c_{L_1}^S, \cdots, c_{L_N}^S) = \sum_{L_i, L_j \in \{L_1, \cdots, L_N\} | i > j} \varepsilon(D_{L_i L_j}, \hat{D}_{L_i L_j}) \quad (4)$$

网络中普通节点测量自己到基准节点之间的距离，然后再通过最小化测量距离和计算距离之间的误差来确定自己相对于基准节点的坐标，其优化目标函数如式(5)所示。通过单纯型下降(simplex downhill)算法对这 2 个目标函数进行求解，进而为网络中每个节点确定一个在嵌入欧氏空间中的坐

标值。

$$f_{\text{obj}2}(c_H^S) = \sum_{L_i \in \{L_1, \dots, L_N\}} \varepsilon(D_{HL_i}, \hat{D}_{HL_i}) \quad (5)$$

GNP 算法是很多网络坐标系统的基础, 如 HNDP^[16], NetPharos^[17], HNPS^[39], 等分层坐标系统就是在将整个网络各子系统中使用 GNP 来进行嵌入的。Ng^[6]随后提出了 NPS(network positioning system)对 GNP 进行改进, 包括: 1) 计算基准节点的坐标值时采用分散方式进行; 2) 层次结构: 一部分节点基于原始基准节点计算自己的坐标值, 其他形成第 1 层。随后第 2 层的节点基于第 1 层的节点坐标计算自身坐标, 同时又作为下一层次节点的基准节点, 按照这种结构依次进行下去。尽管 NPS 在很大程度上优于 GNP, 同时面对基准节点的失效问题也更加健壮, 但由于其采用了逐次迭代的计算方式, 初始计算时一旦出现误差, 则会造成误差的累积放大效应。在实现时为了降低该效应的影响, NPS 只采用了 3 层结构。基于复合空间模型的 Lighthouse^[9]在局部空间也使用 GNP 算法, PIC^[9]也采用了类似的方法进行嵌入, 只是引入了较为复杂的动态基准节点选择机制来避免 GNP 中存在的单点失效问题以及恶意节点的攻击。

3.2 基于模拟的嵌入算法

基于模拟的空间嵌入算法, 其他把网络节点嵌入过程看作一个物理过程, 通过节点相互力的作用最终达到一种平衡, 而此时节点所在位置的坐标就是节点在嵌入空间的坐标。典型的坐标系统有 Vivaldi、BBS、PCoord^[40]等, 下面主要介绍 Vivaldi 和 BBS 2 种较为著名的算法。

节点坐标嵌入的过程本质上就是最小化全局误差的过程, Dabek 认为这一全局最小化过程与物理上通过调整端点位置最小化弹簧弹性势能非常类似, 因此他们将距离预测误差之和最小化问题模拟为弹簧弹性势能最小化问题, 并提出了一种完全分布式距离预测机制 Vivaldi^[2]。节点在加入系统时随机测量到系统中一组节点的距离, 然后根据测量结果确定自己的初始坐标值。每个节点 H_i 与其邻居节点 H_j 都通过一个虚拟弹簧相连, 弹簧的原始长度是 H_i 与 H_j 之间的测量距离 D_{H_i, H_j} , 而当前长度是通过节点当前坐标计算出的距离预测值 \hat{D}_{H_i, H_j} , 因此根据 2 者关系, 弹簧将相应的伸长或压缩, 最终弹

簧在系统预测误差最小时达到一个平衡状态, 此时节点坐标即为最优嵌入坐标。设 \bar{C}_{H_i} 与 \bar{C}_{H_j} 分别代表节点 H_i 与 H_j 的坐标值, 则每经过一次迭代, 节点坐标更新过程如式(6)所示, 其中, δ 代表局部计算步长, $u(\bar{C}_{H_i} - \bar{C}_{H_j})$ 代表弹簧力的作用方向。

$$\bar{C}_{H_i} = \bar{C}_{H_i} + \delta(D_{H_i, H_j} - \hat{D}_{H_i, H_j})u(\bar{C}_{H_i} - \bar{C}_{H_j}) \quad (6)$$

由于采用分布式计算, Vivaldi 算法并不基于全局误差判断系统是否达到平衡, 因而使用基于节点观测到的局部误差设置的局部计算步长 δ 。如果误差很大, δ 就取较大的值以便节点能够迅速移动到一个近似正确的位置, 否则 δ 取较小的值以便节点能够较准确地到达最终位置。此外, Dabek 等人还发现欧氏空间嵌入中的很大一部分误差由接入链路产生, 而位于核心网络中的节点通常具有较小的嵌入误差。因此 Vivaldi 在欧氏坐标中引入了高度参数, 使用一个具有高度参数的 2 维欧氏坐标对时延空间建模。2 个节点间的距离等于其骨干网络中的欧氏距离加上各自的高度值, 进一步提高了预测精度。

相对于 GNP 算法, Vivaldi 的精度较高, 因而许多坐标系统都是基于该算法, 或者对其进行改进。Htrae^[3]通过引入地理距离与时延的启发式算法来加速实际部署网络坐标系统的收敛速度, LAT^[15]通过计算局部误差来代替 Vivaldi 中的高度向量, DS^[41]则通过 Vivaldi 算法和一些全局和局部的时延调整算法来合成与原始数据集具有相同性质的大规模数据集。此外, 一些层次空间模型在局部坐标嵌入时也使用了 Vivaldi 算法, 如 HA(a hierarchical approach)^[12]、Pharos^[13]和 Tread^[42]等。

Y. Shavitt^[37]提出了 BBS, 把网络节点建模一组物理粒子集合, 这些粒子在势能作用下根据牛顿力学原理在欧氏空间中运动, 最终在节点之间的力达到平衡时势能最小, 从而获得其最佳坐标。为了避免节点位置出现振荡, BBS 中又引入摩擦力的概念, 节点在每次运动中都会有一部分能量因为摩擦力作用而消失, 最终达到稳定状态。这种动力学模型能够帮助节点跳出局部最优值, 因此, BBS 在预测精度优于包括 GNP 在内的多种距离预测机制。BBS 不仅适用于欧氏空间, 还适用于双曲空间。但是, BBS 全部计算需要由一个中心节点负责完成, 难以构造分布式实现方式, 导致其扩展性相对较差。PCoord^[40]也采用了类似

BBS 的思想，不再赘述。

3.3 基于数学计算的嵌入算法

在基于数学计算的嵌入算法中，基于矩阵分解的嵌入算法是最为主要的一类，它的提出是为了消除时延空间中存在的非对称性和 TIV 现象。其原理是利用时延空间的低维特性将其嵌入到一个低维的向量空间中，采用一个向量对来表示节点的坐标。典型的模型有虚拟基准节点(virtual landmarks)^[7]、ICS(Internet coordinate system)^[8]、IDES^[11]、Phoenix^[43]、DMF^[44]等。基于矩阵分解的网络坐标系其基本步骤包括以下几点。

1) 通过 PCA 等维数分析工具来判定时延数据集的维数来确定嵌入维数 m 。

2) 在网络中部署 N 个基准节点形成基准节点集，基准节点之间互相测量到达彼此的距离形成一个 $N \times N$ 的距离矩阵 D ，然后由某一中心节点对 D 进行奇异值分解(SVD, singular value decomposition)， $D=UWV^T$ ，那么，距离矩阵 D 中每一个元素可以表示为

$$D_{ij} = \sum_{k=1}^N w_k U_{ik} V_{jk} \quad 1 \leq i \leq N, 1 \leq j \leq N \quad (7)$$

其中， w_i 按降序排列，取其前 m 项($m < N$)对 D_{ij} 进行近似，这样，矩阵 U 的前 m 列组成变换矩阵 U_m ，距离矩阵 D 中任意一个距离向量 D_i 通过变换 $D'_i = U_m^T D_i$ 映射到 m 维向量 D'_i ，该向量即构成基准节点 L_i 在嵌入空间内的坐标值。

3) 普通节点 H_i 加入系统时，首先测量到所有基准节点的距离形成一个距离向量 $V_i = (D_{H_i L_1}, \dots, D_{H_i L_N})$ ，然后根据变换矩阵 U_m 将该向量映射到 m 维空间中，映射后的向量值即节点 H_i 坐标。

IDES、Phoenix、DMF 等坐标系采用了类似的思想，不同是其他采用 NMF (nonnegative matrix factorization) 来对时延矩阵进行分解。与 SVD 不同的是，NMF 没有唯一解，其精度也取决于参数的选择。与 GNP 相比，Virtual Landmarks 和 ICS 在坐标嵌入之前进行了优化，因此能够基于网络特征选择合适的嵌入空间维数，进而提高预测精度。但是其他都需要通过中心节点完成上述优化过程，导致系统容易出现单点故障问题，同时采用固定基准节点又会出现因基准节点失效而造成系统无法正常工作的问题，这些都限制了其在实际大规模网络中的应用。

除了上述基于矩阵的嵌入算法外，Sequoia 提出了一种基于树型空间的嵌入算法。其思想如下，首先选择一个节点作为根节点，然后将其他节点一个个加入到树中，如图 2 所示。一个主机在树中的位置由根节点和另外一个称为锚节点(anchor)的主机共同决定。设待嵌入节点为 B ，根节点为 R ，其锚节点为 A ，在嵌入过程中引入一个虚拟节点 s 将节点 A 、 R 、 B 连接起来，虚拟节点 s 到根节点的距离定义为式(8)。锚节点的选择原则是使得 $d(s, R)$ 最大，即 $\text{Max}\{d(s, R) | A \in D\}$ 。在进行预测时，只需要查找 2 个节点位置，把其他之间虚拟链路的时延叠加起来即可。Sequoia 考虑到负载和准确性，一般将节点划分为多个集合分进行嵌入。

$$d(s, R) = 0.5d(A, R) + d(B, R)d(A, B) \quad (8)$$

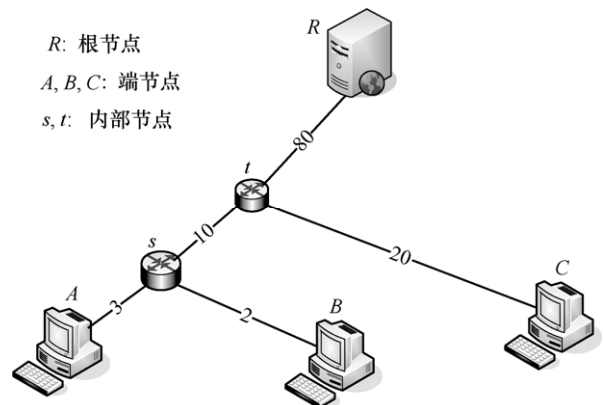


图 2 Sequoia 构建的预测树

综上所述，单空间模型的建模思想比较简单直观，但非常有效。基于单空间的建模使得人们探索了各种可能的数学模型以及嵌入算法。这些模型普遍采用了一种基于基准节点的嵌入算法。相对于直接测量获取 N 个网络节点两两之间的距离需要 $O(N^2)$ 次测量，通过建模只需 $O(L^2 + LN)$ 次测量即可通过预测获取任意节点间距离，其中， L 代表基准节点数目， N 代表普通节点数目，一般情况下 $L \ll N$ 。这些坐标系统的缺陷在其他往往适用于特定的数据集或较小规模的数据集，缺乏通用性。这是由于因特网时延空间的各个部分性质不同，特别是伴随着数据集规模的增大其维数是不收敛的^[26]。

4 复合空间模型

复合空间模型采用了还原法思想，它将网络按其结构分为接入网络 and 核心网络 2 个部分。在这种因特网时延空间模型中，节点间的时延都被划分为 3

个部分:节点到其接入点的时延,主干网络中的时延和另一个节点到其接入点的时延。其中典型的网路坐标系统有 Virtual Landmarks^[7]、Lighthouse^[9]和 NetPharos^[17]等。其他分别建立多个局部空间模型和一个核心空间模型,通过一个全局映射机制来预测端到端的时延。

在 Virtual Landmarks 和 Lighthouse 模型中,有 2 个基本概念:局部坐标(local base)和全局变换矩阵(transition matrix)。一个新节点加入系统后,首先从已经具有位置信息的节点集中随机选择一组节点(称为 lighthouse)作为参考节点,将这些参考节点两两之间组成的向量作为局部基构建一个局部坐标系统,并测量到这些参考节点的距离,然后通过 Gram-Schmidt 变换计算出自己在该局部坐标系统中的位置信息。最后根据全局变换矩阵进行坐标空间映射,将局部坐标变换为全局坐标。

在 NetPharos 中,网络节点按照所处的 AS 和连接关系被划分到边缘预测域和核心预测域中,边缘预测域通过双栈节点(同时位于边缘预测域和核心预测域中)和核心预测域相连,如图 3 所示。对于网络中任意 2 个节点 H_1 和 H_2 ,若其对应的虚拟节点分别为 V_1 和 V_2 ,则节点 H_1 与 H_2 之间距离通过式(9)来进行计算,其中, $D(H_1, V_1)$, $D(V_1, V_2)$, $D(H_2, V_2)$ 表示分别 2 个边缘预测和核心预测中的预测距离。在边缘预测和核心预测域中,NetPharos 使用了 GNP 算法来分别进行预测。此外,文献[17]还证明了采用不同预测域距离信息的迭加不会增加距离预测的相对误差。然而这个证明基于这样一个前提,即分域预测的误差不会大于整个因特网时延空间的预测误差。然而,文献[13~14]都已指出当节点间距离较长时采用欧氏空间更为合适,而在边缘预测域或节点距离较小时则采用具有一定曲率的双曲空间更为合适。因此,并不能保证对于多数节点边缘预测域的误差比整个因特网时延空间的误差要小。

$$D(H_1, H_2) = D(H_1, V_1) + D(V_1, V_2) + D(H_2, V_2) \quad (9)$$

复合空间模型是出现在单空间模型后的一类空间模型,这种模型一般从网络的拓扑结构出发将网络分为中心和边缘 2 部分。然而基于简单的结构划分并没有理论的依据,网络的结构特点并不能直接在时延空间中得到体现。特别是,一些时延性质是从端到端测量获得的,分段研究破坏了这些性质。此外,分段叠加的方式忽略了网络选路的本质,

从一个接入点进入到另一个接入的分组可能经历的路由差异很大。通过对复合空间模型的研究,人们获得了对因特网时延空间独特性质的认识,为提出更加精细的模型做了准备。

GL: 全局基准节点
LL: 局部基准节点

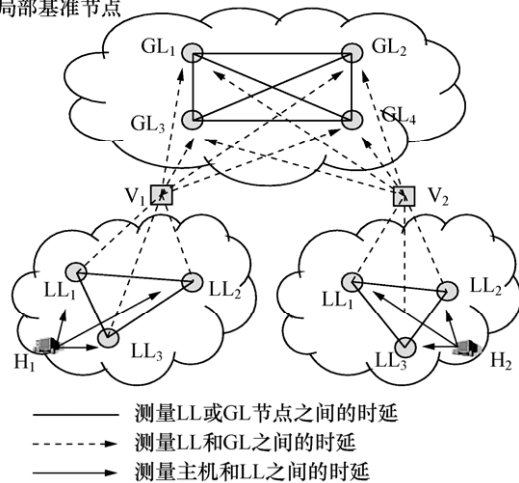


图3 NetPharos的基本架构

5 层次空间型

层次空间模型依据是网络各个部分在不同的位置有着不同的性质。文献[14]当采用基于双曲空间的网络坐标系统预测节点间的时延时,如果节点间的时延小于 100ms 则较为准确,当节点间的时延大于 100ms 时则预测值偏小,采用欧氏空间坐标则较为准确。文献[13]也发现节点距离较大时使用欧氏空间表示更为准确。为了能够较为准确地表示因特网时延空间,层次空间模型将整个因特网时延空间划分为多个子空间 $D = \{D_1, D_2, \dots, D_n\}$, 划分后将每个子空间进行分别嵌入。为了预测不同子空间内节点间的时延,其他为每个节点建立一个全局映射。这样进行时延预测时,采用一种自底向上方式(bottom-up fashion)的预测方法。对于位于同一个嵌入空间的节点,可以用同一个子空间内部的函数来预测 $Dis(i, j, D_i)$, 否则使用全局坐标进行计算 $Dis(i, j, D)$ 。与复合空间模型的区别是,层次空间模型在同样的时延尺度上采用不同坐标系统,而不是在几个子空间中距离的叠加。

$$d'(i, j) = \begin{cases} Dis(i, j, D_i), & i, j \in D_i \\ Dis(i, j, D), & i, j \notin D_i \end{cases} \quad (10)$$

典型的层次空间模型有 HE^[14]、HA(a hierarchical approach)^[12]、Pharos^[13]、Sequoia^[31]和 Toread^[42]

等。其中, HA、Pharos 和 Sequoia 采用了相同的嵌入空间, HE 则使用了不同的嵌入空间。首先介绍这种空间模型的思想, 然后分析其他之间的差异。

层次空间模型同样可以分为以下 2 步。

第 1 步, 将所有的节点按照节点间距离或者连接关系被划分为不同的层次, 一个节点可以在一个或多个簇中。

第 2 步, 为每个节点在不同的层次上分配坐标, 由于一个节点可能在不同层次上加入到多个簇中而拥有多个坐标。

如图 4 所示, 首先在第 0 层包含了所有的节点, 为每个节点分配一个全局坐标 H_0 , 在第 1 层将其划分为 3 个簇, 并为每个节点计算其第 1 层的坐标。在所有的节点中, V_1 和 V_2 同时位于 2 个簇中, 因此其他在第 1 层中具有 2 个坐标, 加上在第 0 层的坐标, 其他共具有 3 个坐标。在预测节点间的时延时, 对于位于同一个簇内的节点, 直接计算第 1 层的坐标获得; 对于不在同一个簇内的节点, 则通过第 0 层的全局坐标计算获得。层次的数目和每个层次上簇的数目取决于数据集自身的性质和所选的聚类算法。

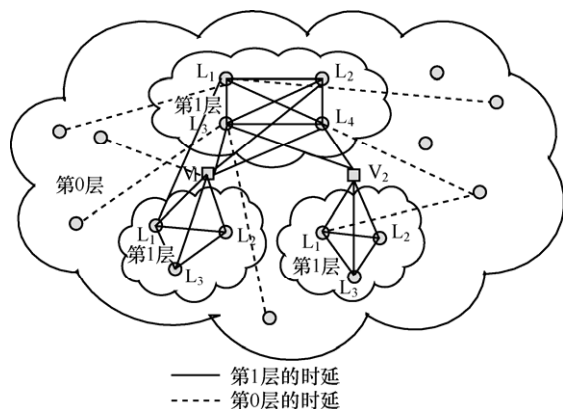


图 4 基于时延距离的节点聚类

在层次空间模型中, 2 个关键技术就是因特网时延空间的划分和子空间的嵌入。子空间的嵌入已经在单空间模型中获得较好地解决, 因此因特网时延空间的层次空间模型主要解决因特网时延空间的划分问题。HA 中子空间的划分层次由人为指定, 具有一定的灵活性; Pharos 则将其直接定义为 2 层, 每一层中采用 Vivaldi 的算法来计算节点坐标; HE 分析了节点时延大小和嵌入空间模型的关系, 发现当节点间时延大于 100ms 时采

用欧氏空间模型更为合适, 小于 100ms 时则采用双曲空间更为合适。Toread^[42]也是根据时延作为阈值来划分子空间的, 不同的是它企图减少子空间内的 TIV 比例。与前者不同的是, Sequoia^[13]采用树型空间进行嵌入, 按照节点所在的区域划分到多个子树中, 不在同一个树内的节点则必须构建全局树来进行预测。

层次空间模型是相对于单空间模型和复合空间模型而提出的, 更多地体现了时延空间的不均匀性。它的优势在于通过时延空间划分, 可以将每个子空间采用更加合适的时延模型进行表示, 因而精度得到了进一步提高。但是这些模型仍然没有解决因特网时延空间建模中面临的一些固有问题, 如 TIV、基准节点的选择等。

6 因特网时延空间建模的难点

6.1 嵌入模型的设计

嵌入模型是保证嵌入精度的首要因素。因特网本身由多种不同类型的物理网络所组成, 而且这些物理网络其性能差异也非常大, 难以通过一个统一的模型对其进行描述。目前虽然已有欧式空间、双曲空间、树型空间等基本嵌入模型以及在其他基础上建立的复合空间和层次空间等模型用于描述因特网时延空间, 但是这些模型都不够精确, 因为这些模型都没有体现因特网时延空间的本质特征。复合模型虽然考虑了网络边缘与核心的差异, 但是网络核心的路由规则复杂, 不是简单叠加能够解决的。层次模型基于时延的划分方式不能保证划分后子集的低维特性, 而数据集的低维特性是进行空间嵌入的依据。文献[26]分析了时延空间的维数特征, 发现因特网时延空间具有分形特征, 因此伴随着因特网规模的增大网络的维数无法用线性维数描述。因此, 在设计时延空间模型时, 必须考虑其扩展性和划分后子空间具有低维特性。

6.2 基准节点的选择

多数的时延空间嵌入算法都分为 2 部分, 即首先选择部分节点作为基准节点进行嵌入, 然后再将其余的节点嵌入到基准节点确定的空间中。因此, 基准节点的选择十分关键, 如果选择的基准节点所在的空间与数据集所在最佳嵌入空间一致, 则可以获得最好的嵌入精度。PIC 研究了基准节点的选择策略, 这些策略包括: 1) 最近节点选择法; 2) 随机节点选择法; 3) 最近选择与随机选择相结合。研究

表明策略 3) 具有最好预测性能。此外, IDES 中也对随机选择策略、K 均值、谱聚类、最大距离(maximum distance)法等 4 种节点选择策略进行了比较。L. Tang^[45]研究了随机、贪婪、K 均值、最大距离的基准节点选择算法和基准节点的数目关系,发现当采用较好的基准节点选择算法时,可以选择较少的基准节点,而当基准节点数目较大时采用随机选择算法即可。但是这些策略不能保证 2 个基准节点所在的空间与最佳嵌入空间一致。因此,如何缺乏全局知识的情况下选择最好的基准节点集,进而确定最佳的嵌入空间值得深入研究。

6.3 TIV 的克服

TIV 作为因特网时延空间的一个基本性质,广泛地、持久地存在于因特网中,因而如何有效地避免 TIV 的影响成为坐标系统必须解决的难题。文献[24]发现路由策略的非高效性是导致 TIV 的重要原因,然而整个因特网的路由策略又是不可获知的。TIV 影响网络坐标系统的精度主要是通过改变数据的维数来影响嵌入精度。

下面以经典的空间嵌入算法 MDS 来分析度量空间嵌入算法引入的误差。设 D 为度量空间, MDS 算法分为 3 步: 1) 获得节点间平方距离矩阵 $D^{(2)} = [d_{ij}^2]$; 2) 对 $D^{(2)}$ 进行双中心化, $B_D = -(1/2)JD^{(2)}J$, 其中, $J = I - n^{-1}11^T$; 3) B_D 进行分解 $B_D = Q Q^T$, 其中, λ_i 为特征值的降序排列。由于 D 为度量空间, 特征值均为非负, 选取其特征值大于 0 的前 m 维向量构成的特征值矩阵为 Λ_m , 则可以 0 误差将 D 嵌入到 m 维欧氏空间中。但是如果数据集 D 中, 存在 TIV 现象则会引入负的特征值 λ_i , 根据矩阵扰动理论可知其引入的误差的下限表示为

$$\sum_{\{i: \lambda_i < 0\}} \lambda_i^2 < e^2 \quad (11)$$

由此可知, 如数据中的 TIV 数目越多, 程度越严重则 λ_i 越大, 引入的嵌入误差 e 越大。为了克服 TIV 的影响, 一些工作研究了 TIV 的性质和判断方法。文献[25]分析了 TIV 的生存期, 发现 80% 的长生命 TIV(long-lived TIV)生存期小于 5h, 同时指出 TIV 的比例会受到处理数据方式的影响。文献[46]提出了一种基于决策树的有监督学习方法来判定 TIV 的发生, 其原理是将时延系统的预测值与实际测量值的统计量作为输入样本, 通过标记的 TIV 来训练决策树, 最后给出一颗

TIV 判定树。

根据上述关于 TIV 的发现, 一些网络坐标采取了一些措施来避免 TIV 的影响以提高因特网时延空间模型的精度和健壮性。文献[41]分析了因特网时延空间的聚簇(cluster)特性, 指出 TIV 在不同的数据簇之间比较严重, 而在簇内则比较轻。文献[47]利用时延较小时 TIV 严重性较轻这一现象, 提出了一种基于时延阈值分层的 Vivaldi 因特网时延空间模型。文献[42]同样利用因特网时延空间 TIV 的聚簇特性, 提出了一种双层因特网时延空间模型。文献[48]进一步分析了时延大小和 TIV 的关系, 在一般情况下时延越大则 TIV 越严重, 但是这种关系却并不总是成立。文献[49]提出了一种基于空间的修复的方法来消除 TIV 的影响, 但是该方法必须要求数据集本身较为完备。这些方法还存在着很多不足, 如何更好地解决 TIV 引入的误差是因特网时延空间建模的一个难点。

7 因特网时延空间建模的主要研究方向

7.1 因特网时延空间模型综合比较

在上文对因特网时延空间模型的基本思想进行了分析, 本节结合时延模型其他几个方面进行了综合分析。

测量开销基于因特网时延空间模型的预测可以减少网络测量开销, 因此建模中使用的测量开销是一个重要方面。

部署方式在时延建模的过程中, 算法的设计确定了网络坐标系统的部署方式。采用集中式部署借助于服务器能够提高系统的收敛速度, 但是不易于扩展; 而分布式系统则不要基础设施的支持, 具有较好的扩展性。

安全性及健壮性距离预测机制必须能够处理大规模网络环境下的多种异常, 尤其是在网络中存在恶意节点或者攻击的情况下, 系统要保证仍能够正常工作, 不能出现单点失效的问题。

从整体上看, 因特网时延空间的建模呈现出以下趋势。

1) 从单一模型向复杂模型转变。单一的空间模型无法反映因特网时延空间中各部分性质, 通过复杂的空间模型, 或者多个几何空间模型的复合模型来对因特网时延空间建模, 成为当前因特网时延空间建模的方向。

2) 模型更加反映了网络的拓扑结构。层次空间

模型或者复合模型的出现,使得模型中体现了更多的网络信息,网络的拓扑、路由信息在模型中得到了体现。

3) 模型更加完善。随着研究的深入,在一个模型中不仅考虑到了时延的动态性,还考虑到 TIV、基准节点和系统部署等多方面的问题。

7.2 因特网时延空间建模的主要研究方向

从因特网时延空间建模的发展过程来看,因特网时延空间的模型日益多样化和复杂化,并逐步地与网络应用相结合。未来的研究方向,主要包括以下5个方面。

1) 因特网时延空间的测量

对因特网时延空间的测量是获取其基本性质的主要途径,也是提出新的因特网时延空间模型的依据。现有的因特网时延性质多是在对 King、Harvard、Meridian、PlanetLab 等少数数据集分析得出的结论。这些数据集规模最大的包括 2 000~3 000 个节点,小的仅包括几百个节点,对于拥有数亿节点的网络来说微不足道,也无法全面反映其性质。因此,如何巧妙地设计实验,对因特网时延空间进行更大规模的测量,对于全面认识因特网时延空间十分有意义。此外,最近测量发现网络结构正在发生新的变化,因特网正由一个具有较高异配系数中心化的网络向同配系数较高的对等网络进行演化^[50]。网络结构的变化是否会引起时延空间性质的变化也需要通过实际测量加以验证。

2) 更加完善的时延空间模型

从最初的空间模型,到后来的层次空间模型,这些模型越来越能够体现时延空间的本质属性。现有模型存在的不可扩展性,以及时延空间中的 TIV 现象都为时延模型在更大范围上的应用带来困难。特别是近些年来,研究主要围绕 TIV 的性质和克服来展开。然而现有的时延空间模型,无论是单空间模型、复合空间模型还是分层模型,在局部往往使用的仍然是度量空间模型,仍然存在着机理性的问题。此外,基准节点选择、网络坐标系统安全性等问题的解决都需要提出全新的模型。

3) 系统的稳定性与安全性

目前,大多数的网络时延模型都是静态的,很多算法没有考虑时延的变化和更新机制。根据矩阵的扰动理论可知,由于时延空间中存在 TIV 现象使

得时延坐标系统为一个不稳定系统,一个节点时延的变化会导致整个系统的大幅扰动甚至失效^[15]。只有设计良好的坐标更新机制,才能保证系统的可用性。此外,很多坐标系统,如 GNP、IDES、Pharos 等为设计的简便使用了固定的基准节点或集中式结构,在面临恶意攻击或网络故障时将会出现问题。虽然有部分研究提出了一些安全措施^[51~53],但都还处于起步阶段。

4) 系统的部署应用

随着因特网时延空间模型的日益精确,网络坐标系统以及一些时延空间性质可以用于辅助提高网络应用的服务质量。如网络坐标系统被用来辅助 SIP 协议构建消息分发树^[54],构建更加高效覆盖网络以缩短节点搜索时间和通信时延^[28,55~58]。同时,许多未来互联研究项目 GENI、FIRE 等中也包含了对网络性能感知的研究^[59,60]。由此可见,将因特网时延空间建模的理论成果付诸于实践,将是一个新的研究方向。在这一过程中,需要考虑到系统的一致性、可扩展性、稳定性、收敛性、安全性、精确性以及存储、计算、通信开销等一系列问题。上述问题,却是理论研究中很少涉及的或没有全面考虑的。

5) 利用网络时延进行 IP 定位

近年来,有越来越多的网络应用需要准确定位用户的地理位置,如广告精准投放(targeted advertisements),确定非法内容的位置以及寻找实施网络攻击的主机。在进行主机定位时,利用的就是时延和地理距离的线性关系,节点间时延的三角形约束关系,典型系统有 CBG、Spotter、Octant。通过研究时延空间中的 TIV 规律可以帮助网络定位算法来提高定位的精度^[61~64],这一研究也十分活跃。

8 结束语

因特网时延空间建模的研究大大地促进了网络测量和网络科学的发展。由于因特网时延空间建模的需要,很多大型的网络测量平台、测量项目得到实施和应用,如 PlanetLab、iPlane、Navigator、Meridian、King。根据因特网时延空间建模的发展阶段,现有的时延模型分为单空间模型、复合空间模型和层次空间模型。这些模型反映了人类对因特网时延空间的认知日益深入。在建模过程中,许多数学方法得到了应用,一些

专门算法被提出来用于坐标嵌入, 这些方法还可以被应用到其他领域。但是, 目前的研究还普遍存在难以适应网络动态变化、对因特网时延空间的性质把握不够全面的问题, 没有哪一种现有的时延模型能够成为公认标准。伴随着网络结构的演化, 因特网时延空间的性质也表现出一些新的特点, 只有深入地把握这些特征才能建立完善准确的模型。总之, 尽管目前有关因特网时延空间建模的研究已经取得了长足的进步, 但这一领域仍有许多问题需要进一步研究。

参考文献:

- [1] 邢长友, 陈鸣. 网络距离预测技术[J]. 软件学报, 2009, 20(9): 2470-2482.
XING C Y, CHEN M. Techniques of network distance prediction[J]. Journal of Software, 2009, 20(9): 2470-2482.
- [2] DABEK F, COX R, KAASHOEK F, *et al.* Vivaldi: a decentralized network coordinate system[J]. SIGCOMM Comput Commun Rev, 2004, 34(4): 15-26.
- [3] AGARWAL S, LORCH J R. Matchmaking for online games and other latency-sensitive P2P systems[A]. Proc of SIGCOMM'09[C]. BARCELONA, SPAIN, 2009. 315-326.
- [4] HOTZ S M. Routing Information Organization to Support Scalable Interdomain Routing with Heterogeneous Path Requirements [D]. Los Angeles, USA: University of Southern California, 1994.
- [5] NG T S E, ZHANG H. Predicting Internet network distance with coordinates-based approaches[A]. Proc of INFOCOM'02[C]. New York, USA, 2002. 170-179.
- [6] NG T S E, ZHANG H. A network positioning system for the Internet[A]. Proc of the USENIX Annual Technical Conf[C]. Boston, Massachusetts, USA, 2004.
- [7] TANG L, CROVELLA M. Virtual landmarks for the Internet[A]. Proc of IMC'03[C]. Miami, Florida, USA, 2003. 143-152.
- [8] LIM H, HOU J, CHOI C. Constructing an Internet coordinate system based on delay measurement[J]. IEEE/ACM Transactions on Networking, 2005, 13(3): 513-525.
- [9] PIAS M, CROWCROFT J, WILBUR S, *et al.* Lighthouse for scalable distributed location[A]. Proc of IPTPS 2003[C]. Berlin, 2003. 278-291.
- [10] COSTA M, ROWSTRON A, ROWSTRON A, *et al.* PIC: practical Internet coordinates for distance estimation[A]. Proc of ICDCS 2004[C]. Washington, DC, USA, 2004. 178-187.
- [11] MAO Y, SAUL L K. IDES: an internet distance estimation service for large networks[J]. IEEE Journal on Selected Areas in Communications, 2006, 24: 2273-2284.
- [12] ZHANG R, HU Y C, LIN X, *et al.* A hierarchical approach to Internet distance prediction stable and accurate network coordinates[A]. Proc of ICDCS2006[C]. Lisboa, Portugal, 2006.
- [13] CHEN Y, XIONG Y, SHI X, *et al.* Pharos: accurate and decentralized network coordinate system[A]. Proc of IEEE GLOBECOM 2007[C]. Washington, DC, 2007. 539 - 548.
- [14] LUMEZANU C, SPRING N. Measurement manipulation and space selection in network coordinates[A]. Proc of IEEE ICDCS[C]. Beijing, China, 2008. 361-368.
- [15] LEE S, ZHANG Z L, SAHU S, *et al.* On suitability of euclidean embedding for host-based network coordinate systems[J]. IEEE Transactions on Networking, 2010, 18(1): 27-40.
- [16] XING C, CHEN M. HNDP: a novel network distance prediction mechanism[A]. Proc of the IFIP NPC[C]. Dalian, China, 2007.
- [17] XING C, CHEN M. A virtual node based network distance prediction mechanism[A]. Proc of GlobeCom2008[C]. New Orleans, LA, USA, 2008. 1-5.
- [18] LEDLIE J, GARDNER P, SELTZER M. Network coordinates in the wild[A]. Proc of NSDI 2007[C]. Cambridge, MA, 2007.
- [19] LEONARD D, LOGUINOV D. Turbo king: framework for large-scale internet delay measurements[A]. Proc of the IEEE INFOCOM[C]. Phoenix, AZ, USA, 2008.
- [20] One way ping[EB/OL]. <http://e2epi.internet2.edu/owamp/>, 2012.
- [21] IETF. IP performance metrics (IPPM) charter[EB/OL]. <http://www.ietf.org/mail-archive/web/ippm/>, 2012.
- [22] SHAVITT Y, SHIR E. DIMES: let the Internet measure itself[J]. ACM SIGCOMM Computer Communication Review, 2005, 35(5): 71-74.
- [23] SHAVITT Y, TANKEL T. On the curvature of the Internet and its usage for overlay construction and distance estimation[A]. Proc of INFOCOM 2004[C]. 2004.
- [24] ZHENG H, LUA E K, PIAS M, *et al.* Internet routing policies and round-trip-times[A]. Proc of PAM2005[C]. Boston, MA, USA, 2005. 236-250.
- [25] LUMEZANU C, BADEN R, SPRING N, *et al.* Triangle inequality variations in the Internet[A]. Proc of IMC 2009[C]. Chicago, Illinois, USA, 2009. 177-183.
- [26] ABRAHAO B, KLEINBERG R. On the Internet delay space dimensionality[A]. Proc of IMC 2008[C]. Vouliagmeni, Greece, 2008. 157-168.
- [27] ZHANG B, NG T S E. Exploiting internet delay space properties for

- selecting distinct network locations[A]. Proc of INFOCOM'09[C]. Brazil, 2009. 1467-1475.
- [28] CRISTIAN L. Using Internet Geometry to Improve End-to-end Communication Performance [D]. Washington, DC, USA: University of Maryland, 2009.
- [29] BREITBART Y, GAROFALAKIS M, JAI B, *et al.* Topology discovery in heterogeneous IP networks: the netinventory system[J]. IEEE/ACM Trans on Networking, 2004,12(3):401-414.
- [30] HU N, STEENKISTE P. Exploiting internet route sharing for large scale available bandwidth estimation[A]. Proc of IMC 2005[C]. Berkeley, CA, USA, 2005. 187-192.
- [31] VENUGOPALAN, MALKHI D, KUHN F, *et al.* On the treeness of Internet latency and bandwidth[A]. Proc of SIGMETRICS 2009[C]. SEATTLE, WA, USA, 2009. 61-72.
- [32] ALOK S. Efficient Techniques for End-to-end Bandwidth Estimation: Performance Evaluations and Scalable Deployment[D]. Los Angeles, USA: University of North Carolina, 2009.
- [33] XING C, CHEN M, YANG L. Predicting available bandwidth of Internet path with ultra metric space-based approaches[A]. Proc of IEEE GlobeCom 2009[C]. Honolulu, Hawaii, USA, 2009.1-6.
- [34] WANG Y A, HUANG C, LI J, *et al.* Queen: Estimating packet loss rate between arbitrary Internet hosts[A]. Proc of PAM2009[C]. Seoul, South Korea, 2009. 57-66.
- [35] ELSE B, FÖRSCHLER A, FUHRMANN T. Spring for vivaldi -- orchestrating hierarchical network coordinates[A]. Proc of Peer-to-Peer Computing[C]. 2010. 1-4.
- [36] MADHYASTHA H, ISDAL T, PLATEK M, *et al.* Iplane: an information plane for distributed services[A]. Proc of OSDI 2006[C]. SEATTLE, WA, USA, 2006. 367-380.
- [37] SHAVITT Y, TANKEL T. Big-bang simulation for embedding network distances in euclidean space[J]. IEEE/ACM Transactions on Networking, 2004, 12 (6): 993-1006.
- [38] SHAVITT Y, TANKEL T. Hyperbolic embedding of Internet graph for distance estimation and overlay construction[J]. IEEE/ACM Trans Networking, 2008, 16(1):25-36.
- [39] DONG C B, WANG G D, ZHANG X, *et al.* Two-layer network coordinate system for Internet distance prediction[A]. Proc of ICUMT[C]. St.Petersburg, Russia, 2009, 1-6.
- [40] LEHMAN L, LERMAN S. Pcoord: network position estimation using Peer-to-Peer measurements[A]. Proc of the 3rd IEEE Int'l Symp on Network Computing and Applications[C]. Cambridge, MA, USA, 2004. 15-24.
- [41] ZHANG B, Ng T S E, NANDI A, *et al.* Measurement-based analysis, modeling, and synthesis of the Internet delay space[J]. IEEE/ACM Transactions on Networking, 2010, 18:229-242.
- [42] ZHU Y, CHEN Y, ZHANG Z, *et al.* Taming the triangle inequality violations with network coordinate system on real Internet[A]. Proc of CoNEXT'10[C]. Philadelphia, USA. 2010.
- [43] CHEN Y, WANG X, SONG X, *et al.* Phoenix: towards an accurate, practical and decentralized network coordinate system[A]. Proc of IFIP Networking[C]. Aachen, Germany, 2009.
- [44] LIAO Y, GEURTS P, LEDUC G. Network distance prediction based on decentralized matrix factorization[A]. Proc of IFIP Networking[C]. Chennai, India, 2010.
- [45] TANG L, CROVELLA M. Geometric exploration of the landmark selection problem[A]. Proc of PAM2004[C]. France, 2004.
- [46] LIAO Y, MOHAMED A K, GUEYE B, *et al.* Work in Progress: Detecting Triangle Inequality Violations in Internet Coordinate Systems by Supervised Learning[R]. 2009.
- [47] KAAFAR M A, GUEYE B, CANTIN F, *et al.* Towards a two-tier Internet coordinate system to mitigate the impact of triangle inequality violations[A]. Proc of IFIP Networking[C]. Valencia, Spain, 2008. 397-408.
- [48] WANG G, ZHANG B, NG T S E. Towards network triangle inequality violation aware distributed systems[A]. Proc of IMC'07[C]. San Diego, 2007. 175-188.
- [49] 王占丰,陈鸣,邢长友. S-Vivaldi :一种基于空间修复的因特网时延空间嵌入算法[J]. 通信学报, 2012,33(3): 163-170.
- WANG Z F, CHEN M, XING C Y. S-vivaldi: a space repairing based internet delay space embedding algorithm[J]. Journal on Communications, 2012,33(3): 163-170.
- [50] HADDADI H, FAY D, UHLIG S, *et al.* Mixing biases: structural changes in the AS topology evolution[A]. Proc. of TAM[C]. ZURICH, SWITZERLAND, 2010. 32-45.
- [51] KAAFAR M A, MATHY L, BARAKAT C, *et al.* Securing Internet coordinate system: embedding phase[A]. Proc of the ACM SIGCOMM[C]. Kyoto, Japan, 2007.
- [52] KAAFAR M A, MATHY L, TURLETTI T, *et al.* Real attacks on virtual networks: vivaldi out of tune[A]. Proc of the SIGCOMM Workshop on Large-scale Attack Defense[C]. Pisa, Italy, 2006. 139-146.
- [53] KAAFAR M A, MATHY L, TURLETTI T, *et al.* Virtual networks under attack: disrupting internet coordinate systems[A]. Proc of the 2nd CoNext Conf[C]. Lisboa, Portugal, 2006.
- [54] SMITH D A H, HUANG C, YAN J. Hierarchical notification dissemination for IMS presence using network coordinates[A]. Proc of GLOBECOM[C]. Florida, USA, 2010. 1-5.

- [55] KOJIMA T, ASAHARA M, KONO K, *et al.* Embedding network coordinates into the heart of distributed hash tables[A]. Proc of Peer-to-Peer Computing[C]. SEATTLE, WA, USA, 2009, 155-158.
- [56] HU N, WANG D, MA Y, *et al.* Genius: Peer-to-Peer location-aware gossip using network coordinates[J]. Lecture Notes in Computer Science, 2005, 3516:163-170.
- [57] LEE S, APNOMS. Exploiting network distance based euclidean coordinates for the one hop relay selection[J]. Lecture Notes in Computer Science, 2009, 5787:527-530.
- [58] XIE H, YANG Y, KRISHNAMURTHY A, *et al.* P4P: provider portal for applications[A]. Proc of the ACM SIGCOMM 2008[C]. Seattle, WA, USA, 2008. 351-362.
- [59] ELLIOTT C. GENI: opening up new classes of experiments in global networking[J]. IEEE Internet Computing, 2010, 14: 39-42.
- [60] PAUL S. Architectures for the future networks and the next generation internet: a survey[J]. Computer Communications, 2011, 34(1): 2-42.
- [61] LAKI S, MÁTRAY P, HÁGA P. Spotter: a model based active geolocation service[A]. Proc of INFOCOM[C]. Shanghai, China, 2011.
- [62] LAKI S, MÁTRAY P, HÁGA P, *et al.* A model based approach for improving router geolocation[J]. Computer Networks, 2010, 54(9): 1490-1501.
- [63] GUEYE B, ZIVIANI A, CROVELLA M, *et al.* Constraint-based geolocation of Internet hosts[J]. IEEE/ACM Transactions on Networking, 2004, 14(6):1219-1232.
- [64] WONG B, STOYANOV I, SIRER E G. Octant: a comprehensive framework for the geolocalization of Internet hosts[A]. Proc of NSDI 2007[C]. Cambridge, Massachusetts, 2007.

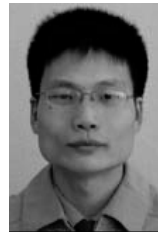
作者简介：



王占丰 (1982-), 男, 河北临城人, 解放军理工大学博士生, 主要研究方向为计算机网络和网络测量。



陈鸣 (1956-), 男, 江苏无锡人, 博士, 解放军理工大学教授、博士生导师, 主要研究方向为网络体系结构、网络管理、网络测量和分布式系统。



邢长友 (1982-), 男, 河南杞县人, 博士, 解放军理工大学讲师, 主要研究方向为计算机网络、网络测量和分布式系统。

李兵 (1967-), 男, 四川金堂人, 硕士, 解放军理工大学副教授, 主要研究方向为计算机网络和数据库应用。

仇小峰 (1976-), 男, 江苏泰州人, 博士, 解放军理工大学讲师, 主要研究方向为网络测量和网络安全。