一种层次化网络距离预测机制

邢长友 陈 鸣
(解放军理工大学指挥自动化学院 南京 210007)

摘 要 网络坐标系统向分布式应用提供了一种高效的网络距离信息获取机制,但现有基于单一度量空间嵌入的距离预测机制难以精确描述因特网复杂的层次结构特征,进而导致较大的距离预测误差.文章提出了一种分域的层次化网络距离预测机制 NetPharos,该机制根据因特网结构以及性能特征将其划分为相互独立的核心预测域和边缘预测域,通过相关边缘预测域和核心预测域内预测值迭加获取节点间距离信息.理论分析和仿真实验显示,NetPharos 能够有效解决预测过程中短距离和长距离的相互干扰问题,提高预测精度.

关键词 网络坐标系统;距离预测;层次化结构;预测域;空间嵌入
中图法分类号 TP393 **DOI 号**: 10.3724/SP.J.1016.2010.00356

A Hierarchical Network Distance Prediction Mechanism

XING Chang-You CHEN Ming
(*Institute of Command Automation, PLA University of Science and Technology, Nanjing* 210007)

Abstract Network Coordinate system provides an efficient mechanism to obtain network distance (latency) information with limited times of measurement. However, prediction mechanisms based on single metric space embedding cannot describe the complex hierarchical structure of Internet precisely, and result in great prediction errors. By analyzing the hierarchical structure feature of Internet, a region-partition based hierarchical network distance prediction mechanism named NetPharos is proposed, which divides Internet into independent core prediction region and many edge prediction regions based on the structure and performance feature of Internet. Distances between any two nodes are obtained by accumulating distances in their edge regions and the core region. Theoretic analysis and simulation results show that NetPharos can avoid the interference between short distances and long distances during distance prediction process, and improve distance prediction accuracy.

Keywords network coordinate system; distance predication; hierarchical structure; predication region; space embedding

1 引 言

因特网服务被定义为尽力而为的分组交付服务,它并不向上层应用提供不同路径的性能信息.然而当前许多大型分布式应用都需要依赖于该信息进行性能优化,导致上述设计理念不能适应这些新型网络应用的需求.例如,结构化 P2P 应用基于性能信息可以构造更加高效的分布式 Hash 表结构[1];在多播系统中,性能信息可以协助构造更加优化的

收稿日期:2007-01-19;最终修改稿收到日期:2008-02-29.本课题得到国家自然科学基金重大研究计划资助项目(90304016)和国家"八六三"高技术研究发展计划项目(2007AA01Z418)资助.**邢长友**,男,1982 年生,博士,讲师,主要研究方向包括网络测量、分布式系统等.E-mail: xcy@plaust.edu.cn. **陈 鸣**,男,1956 年生,博士,教授,博士生导师,主要研究领域为网络测量、网络管理、分布式系统和网络体系结构等.

多播树[2]；基于 CDN 的流媒体服务，可以根据网络性能状况选择最优服务器为用户提供服务[3]；在 Skype 等需要使用中继节点桥接位于 NAT/防火墙后主机节点的 VoIP 系统中，性能信息能够帮助选择具有最优性能的节点作为桥接节点，从而为用户提供最佳通话质量[4]. 最后，ISP 本身也可以通过性能信息分析因特网可达性、选路不稳定性以及检测 DDoS 攻击等.

为了获取所需性能信息，目前不少应用都设计了相应的测量机制. 然而，这种解决方案会造成诸多问题：一方面测量需要花费一定的时间，导致系统无法及时响应用户的性能信息查询请求；另一方面如果大量用户都采用该方式将会严重侵扰网络. 在各种网络性能指标中，网络距离（往返时延）是一个非常重要而又相对容易获取的参数，因此，设计一个合适模型实现对网络距离的有效预测具有重要的现实意义. 根据网络坐标系统的基本理念，本文中我们提出了一种层次化的网络距离预测机制 NetPharos，从而能够基于部分距离测量信息，通过坐标空间嵌入机制实现网络中任意两个节点之间的时延预测，为向因特网中各种应用提供网络距离信息给出了一种解决思路.

本文第 2 节概述网络距离预测相关研究工作；然后第 3 节总结网络坐标系统的基本概念，分析因特网的层次结构，提出并详细描述层次化的距离预测机制 NetPharos；随后第 4 节基于多个性能指标对 NetPharos 进行理论和仿真分析；最后第 5 节总结全文.

2　相关研究工作

Ng 最先提出了一种通过虚拟坐标嵌入实现网络距离预测的机制 GNP（Global Network Positioning）[5]. 在该机制中，将因特网建模为一个欧氏空间，根据测量得到的距离关系为网络中每个节点分配一个坐标值，不同节点间距离基于坐标值采用欧氏空间距离公式计算得出. BBS（Big-Bang Simulation）将网络节点建模为势能作用下根据牛顿力学在欧氏空间中运动的粒子，并且其势能由节点间全部的嵌入误差组成，算法最终在总嵌入误差降到最低时终止，BBS 克服了 GNP 算法收敛速度慢等问题[6]. PIC 允许在构建节点坐标值的过程中动态选择基准节点，相对于使用固定基准节点的 GNP 而言扩展性更好[7]. Vivaldi 通过完全分布式的结构实现坐标嵌入，简化了网络坐标系统的部署[8]. Shavitt 等研究了因特网距离空间的曲率特征，提出负曲率度量空间更适合描述因特网距离，并据此构造了一种基于双曲空间的距离预测机制[9]. ICS（Internet Coordinate System）[10] 虽然同样是基于虚拟坐标的距离预测机制，但是它使用 Lipschitz 空间嵌入，同时使用主成分分析法提取网络拓扑信息，达到降低嵌入空间维数的目的. Zhang 等人提出了一种分层的因特网距离预测机制，每个节点维护多个坐标值，分别用于预测不同范围的距离[11]. 然而，该模型一方面难以对距离范围进行有效划分，另一方面在描述因特网实际结构时仍显得较为粗糙.

3　层次化网络距离预测机制

3.1　网络坐标系统基本概念

度量空间嵌入（metric space embedding）是数学中的一个重要研究内容，它主要关注于在较低扭曲度下实现高维、复杂度量空间到低维度量空间的嵌入. 网络坐标系统（network coordinate system）借鉴了度量空间嵌入的基本思想，通过坐标空间嵌入方式将网络距离空间映射到一个几何空间中，每个网络节点在该几何空间中分配一个坐标值，节点间距离可以根据其坐标值通过距离公式计算得出.

对于一个具有 N 个节点 $H=\{H_1,H_2,H_3,\cdots,H_N\}$ 的网络，令 D_{ij} 代表节点 H_i 到节点 H_j 的距离，并约定节点到自身的距离为 0，则 H_i 到网络中所有节点的距离组成一个 N 维距离向量 $(D_{i1},D_{i2},\cdots,D_{iN})$. 因此，所有节点间的距离就构成一个 $N\times N$ 的非负距离矩阵 $\boldsymbol{D}$

$$\boldsymbol{D}=\begin{pmatrix} D_{11} & D_{12} & \cdots & D_{1N} \\ D_{21} & D_{22} & \cdots & D_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ D_{N1} & D_{N2} & \cdots & D_{NN} \end{pmatrix} \tag{1}$$

网络坐标系统就是设计一个 M 维的虚拟坐标空间，从而能够将上述距离矩阵中每一行 N 维距离向量均嵌入到 M 维空间中，同时使得嵌入后根据坐标值计算得出的距离与实际测量得到的距离误差值最小. 即构造一个网络节点到 M 维向量的映射 f：$H\to R^M$，为网络中每个节点分配一个虚拟坐标值. 这样，任意两个网络节点 H_i 与 H_j 之间距离则表示为从 M 维向量到非负实数的映射 g：$R^M\to R$.

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

其中，$\hat{\boldsymbol{D}}_{ij}$ 代表节点 H_i 和 H_j 之间的预测距离，$f(H_i)$ 是一个 M 维的向量，代表节点 H_i 经过映射后在嵌入空间内的虚拟坐标.

$$f(H_i)=\boldsymbol{H}=(H_{i1},H_{i2},\cdots,H_{iM}) \tag{3}$$

例如，在欧氏空间中，我们有距离预测值

$$\hat{D}_{ij}=\|f(H_i)-f(H_j)\|=\Big(\sum_{k=1}^{M}(H_{ik}-H_{jk})^2\Big)^{1/2} \tag{4}$$

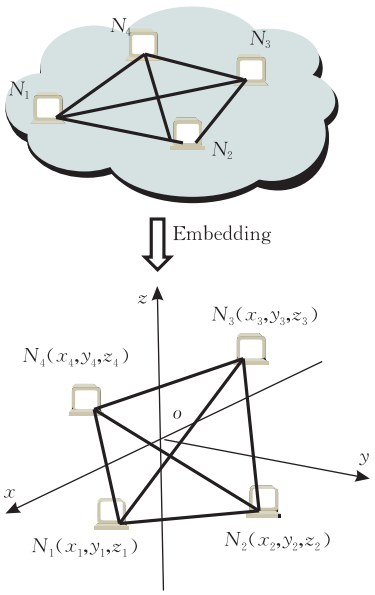


图 1 网络坐标系统原理图

图 1 描述了将一个网络嵌入到三维欧氏空间的示意，每个节点维护自己在嵌入空间内的坐标. 根据节点坐标可以计算得出网络中任意两个节点间距离，避免了直接测量对网络的侵扰，同时相对于原来通过 $N\times N$ 矩阵描述节点间距离，网络坐标系统也简化了距离描述方式.

3.2 因特网层次结构

总体上说，因特网层次结构可以分为国际连接、国家主干、区域网络和本地网. 区域网内的节点彼此密切连接，使得网络内部具有很高的聚合系数，而这些高度聚合的区域网域由国际主干和国家主干相互联系起来[12]. 这样，核心网络是由一些大型路由器通过高速链路连接所组成的网状结构，而边缘网络则由具有高连通度、类似树型结构的区域网络(称之为接入网)组成，边缘网络通过较低速率的链路接入核心网络.

在因特网中，位于网络边缘的接入网络通过分层的 ISP(因特网服务提供商)等级结构与因特网其它部分相连. 如图 2 所示，在该等级结构的最顶层是数量相对较少的第一层 ISP，其链路速率经常是 10Gbps 或更高，并且通常覆盖国际区域. 第 2 层 ISP 通常具有区域性或国家性覆盖规模，并且仅与少数第 1 层 ISP 相连接，它们的主要功能就是引导流量通过它所连接的第 1 层 ISP，同时不同第 2 层 ISP 之间也会具有一些对等连接. 在第 2 层 ISP 之下是较低层的 ISP，这些 ISP 经过一个或多个第 2 层 ISP 与更大的因特网相连. 在该等级结构的底部是接入 ISP，它们通过住宅接入、公司接入和无线接入等方式与本地网相连. 可见，因特网具有典型的层次化结构，并且各层次的性能具有不均匀特性.

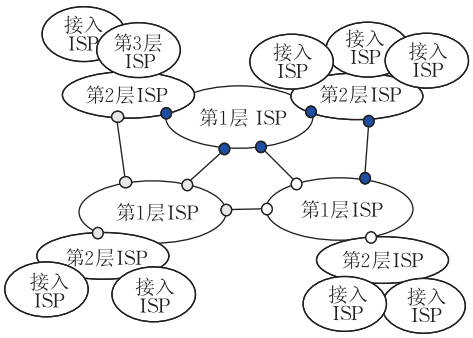


图 2 因特网层次结构

3.3 层次化网络距离预测机制 NetPharos

因特网的层次化结构导致网络性能分区域的不均匀，进而影响其时延分布模式. 在核心网络中，高速路由器通过专用高速链路相连，所以网络时延状况相对稳定；而在边缘网络，由于节点数量很多，网络带宽往往受限，因此网络时延变化范围较大、变化速率较快. 例如，目前广泛使用的 DSL 以及 Cable 因特网接入方式性能特征与核心网络性能特征有很大区别，其中由于排队等因素使得网络节点间时延变化范围具有近 4 个数量级的差别[13]. 并且这种层次化网络结构也导致因特网选路的层次性，例如，当前存在不同的域间以及域内选路协议，使得因特网端到端路径并非时延意义上的最优路径，因此节点间距离会违反三角形不等式的约束[14]，导致无法通过任何一种度量空间对其进行精确描述.

现有的网络距离预测机制都试图通过诸如欧氏空间这类单一度量空间刻画层次化的网络结构，导致在坐标空间嵌入过程中损失了很多状态信息，造成较大的距离预测误差. 另一方面，许多单一空间的距离预测机制都需要使用全网统一的基准节点，又影响了算法的扩展性，而且跨域的性能测量对网络造成的侵扰也较大. 为此，我们设计实现一种分域的层次化网络距离预测机制 NetPharos，根据结构以及性能特征将因特网划分为多个相互独立的预测

域，不同预测域在进行坐标空间嵌入过程中不存在干扰. 通过这样一种结构，既能够更加精确地描述因特网结构特征，又消除了不同 ISP 所属预测域之间的影响，有效解决网络坐标系统在大规模网络中的部署问题.

3.3.1　NetPharos 定义

在 NetPharos 中，根据因特网的结构以及性能特征，将整个因特网划分为边缘网络与核心网络两部分，核心网络作为一个整体构建坐标嵌入空间，边缘网络按照网络距离、所属关系等限制分区域构建独立的坐标空间，不同空间内网络节点坐标值之间相互独立. 同时，可以根据网络距离特征，独立地为每个预测域选择最佳嵌入空间以及预测结果更新频率等. 例如，由于核心网络距离相对稳定，同时节点距离违反三角形不等式等情况很少见，因此欧氏空间就能够很好地实现嵌入. 而考虑到边缘网络通常具有的类树型拓扑，树型空间可能更加适合边缘网络等. 在距离预测过程中，通过“边缘-核心-边缘”度量空间内距离迭加获得网络节点间距离. 为了对该算法进行描述，我们首先给出如下一些基本定义.

定义 1.　预测域 *Region* 指一个或多个 AS 所覆盖的并能进行虚拟坐标空间嵌入的区域，该区域能独立地嵌入到一个自定义的虚拟坐标空间中.

定义 2.　核心预测域 *Core* 指第 1 层和第 2 层 ISP 的 AS 所组成的预测域，该域通常位于网络的核心. 在 NetPharos 中，核心预测域只有 1 个.

定义 3.　边缘预测域 *Edge* 指第 3 层和接入 ISP 的 AS 所组成的预测域，该域通常位于网络的边缘. 不同边缘预测域之间没有交集，每个边缘预测域与核心预测域之间根据接入关系有一个或多个相交节点.

定义 4.　双栈节点. 边缘预测域与核心预测域的相交节点，它们同时参与两个预测域的坐标空间嵌入、具有两个坐标空间的坐标值.

定义 5.　NetPharos 中逻辑约束关系：

$NetPharos=\{Region,RS\}$；

$Region=\{Core,Edge_i\mid i=1,2,\cdots,m\}$；

$Edge=\bigcup_{i=1}^{m}Edge_i$；

$Dual=\{H_i\mid H_i\in Edge\wedge H_i\in Core,\ i=1,2,\cdots,n\}$；

$NE=\{H_i\mid H_i\in Edge\wedge H_i\notin Dual,\ i=1,2,\cdots,n\}$；

$SRS=\{\langle H_i,H_j\rangle\mid\forall H_i\in Edge_k,\forall H_j\in Edge_k\}$；

$MRS=\{\langle H_i,H_j\rangle\mid\forall H_i\in Edge_k,\forall H_j\in Edge_s,k\neq s\}$；

$NDRS=\{\langle N_i,N_j\rangle\mid\forall N_i\in NE,\forall N_j\in Dual\}$；

$DNRS=\{\langle H_i,H_j\rangle\mid\forall H_i\in Dual,\forall H_j\in NE\}$；

$RS=SRS\cup MRS$.

其中，H_i 表示系统中第 i 个节点；$Edge_i$ 代表网络中第 i 个边缘预测域；*Dual* 代表双栈节点集合；*NE* 代表边缘预测域中普通非双栈节点集合；*RS* 是节点间关系集合，*SRS* 是同一个预测域内节点间关系集合，*MRS* 是不同预测域内节点间关系集合；*NDRS* 和 *DNRS* 是普通节点与双栈节点之间的关系集合.

定义 6.　节点逻辑约束关系：

$$SES_j=ran\ NDRS_j(NDRS_j\subseteq NDRS,\ dom\ NDRS_j=\{H_j\})$$

$$SDS_j=ran\ DNRS_j(DNRS_j\subseteq DNRS,\ dom\ DNRS_j=\{D_j\})$$

SES_j 是普通节点 H_j 的双栈节点集合. SDS_j 是双栈节点 D_j 所覆盖的普通节点集合.

NetPharos 的工作原理可参见图 3 中一个简单的网络结构示例. 其中按照拓扑连接关系可将其划分为 5 个预测域，中间一个独立网络结构以及周边双栈节点被定义为核心预测域，每个边缘网络以及该网络相关的双栈节点组成一个独立的边缘预测域. 每个边缘预测域与核心预测域之间存在一个或多个双栈节点.

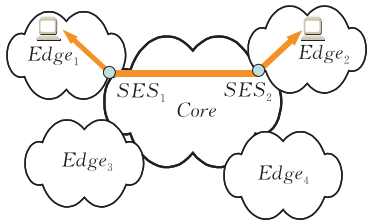


图 3　基于网络层次结构的预测域划分

各个预测域在经过坐标嵌入确定域内所有节点坐标值后，在进行距离预测时，同一个预测域内的节点间距离直接根据节点坐标值计算得出，不同预测域内节点间的距离通过两个节点到各自双栈节点间的距离以及双栈节点之间的距离迭加获得.

```
if (SES_i ∩ SES_j == NULL)
    D(H_i, H_j) = D(H_i, SES_i) + D(SES_i, SES_j) +
                  D(SES_j, H_j);
else
    compute distance directly
    using node coordinate;
```

实际部署时，如果某个边缘网络的双栈节点不止一个，即对节点 H_j 而言，若 $|SES_j|>1$，那么在距离迭加过程中必须为端节点选择最佳双栈节点. 例

如，在图 4 中 $|SES_a|=|SES_b|=3$，$|SES_c|=|SES_d|=2$，并且节点 a_0 在与节点 c_0 和节点 d_0 通信时分别使用了双栈节点 a_3 和 a_5，故而不能采用一个双栈节点代表所有情况，能否正确选择双栈节点直接影响迭加后距离预测的精度. 对此可以采用两种处理策略：

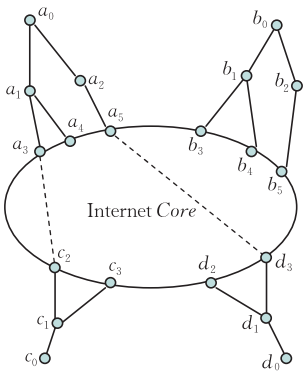


图 4 双栈节点选择

(1) 对所有可能结果进行计算，并选择其中最小值作为节点间距离预测值. 这一策略的优点是计算比较简单，但由于因特网选路过程中并非采用最短距离作为唯一标准，因此会导致一定的预测误差.

(2) 利用 Source Sink Tree 选择最佳双栈节点[15]. Source Sink Tree 是 Hu 等人提出的一种网络抽象结构，其基本思想是在网络中部署一些用于测量的基准节点，每个普通节点通过向基准节点发起 traceroute 测量获得自己的 AS 树. 不同节点之间通过判断规则基于 AS 树获得共有 AS，从而确定通信过程中使用的双栈节点.

Source Sink Tree 能够保证选择的双栈节点更符合因特网实际选路中由边缘网络到核心网络的接入节点，因此预测结果优于简单选择距离最近节点作为双栈节点，但其需要额外耗费进行基准节点部署以及路由信息测量. 然而，由于因特网路由的相对稳定性，节点进行上述测量的频率可以设置为很低，因此利用该机制进行双栈节点选择仍具有很高的应用价值.

3.3.2 NetPharos 算法描述

根据前述讨论，我们给出层次化的网络距离预测算法 NetPharos 如下.

1. 初始化：

a) 根据网络拓扑结构，划分核心预测域 *Core*、边缘预测域 *Edge*.

b) 边缘网络在其到核心网络的入口处部署双栈节点 *Dual*.

c) 在每一个预测域中随机选择部分节点作为基准节点.

2. 计算虚拟坐标：

a) 每个预测域内的基准节点互相测量到达彼此的距离 d，构造在虚拟坐标空间中的坐标值 $(x,y,\cdots,u)$，通过最小化距离测量值与预测值之间的误差确定 $x,y,\cdots,u$ 的值.

b) 预测域中用户节点测量自己到本预测域内基准节点的距离，然后采用类似机制确定节点坐标值. 双栈节点由于同时位于两个预测域中，因此具有两个坐标值.

c) 用户节点 H_j 在本地记录其双栈节点 SES_j 的坐标值，如果双栈节点坐标发生改变，则通知本域内所有节点进行更新.

3. 距离预测：

节点间关系分为 SRS 和 MRS 两类，对于 SRS，距离直接根据节点坐标值计算得出，否则通过两个节点到各自双栈节点间的距离以及双栈节点之间的距离迭加构造. 具体的迭加策略在前面已经进行了详细讨论.

4 NetPharos 预测精度分析

4.1 理论分析

对一种特定网络坐标系统，我们可以从下述典型指标对其距离预测性能进行评价.

相对误差：它描述了距离预测值与实际测量值之间的相对差异情况，是评价一种预测机制优劣的最基本指标. 相对误差包括单向相对误差 *URE* (Unidirectional Relative Error) 和双向相对误差 *BRE* (Bidirectional Relative Error) 两种，其定义分别如式(5)和式(6)所示. *URE* 越小代表预测误差越小，而 $BRE>0$ 表明预测值偏高，否则表明预测值偏低.

$$URE_{ij}=|\hat{D}_{ij}-D_{ij}|/D_{ij} \tag{5}$$

$$BRE_{ij}=(\hat{D}_{ij}-D_{ij})/D_{ij} \tag{6}$$

最近节点误判误差 *CNLS* (Closest Node Loss Significance). 进行距离预测的一个重要目标就是帮助网络应用寻找满足某一条件的距离最近节点，因此，对任意一个节点 H_i 而言，我们希望基于距离预测结果选择得到的最近节点与实际网络中的最近节点为同一节点，或者至少两者到 H_i 的距离要尽可能接近. 最近节点误判误差即是对这一要求进行评价的指标，针对网络中任意一个节点 H_i，其最近节点误判误差定义为[16]

$$CNLS(H_i)=\frac{|D(H_i,H_j)-D(H_i,H_k)|}{D(H_i,H_j)} \tag{7}$$

其中 H_j 是实际网络中与 H_i 最近的节点，H_k 是预测算法得到的到 H_i 的最近节点. $D(H_i,H_j)$，$D(H_i,H_k)$ 分别代表相应节点间的测量距离.

定理. 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.

证明. 对于任意两个预测域中两个节点 H_A 和 H_B，其对应双栈节点分别为 D_A 和 D_B，设 $D(H_i,H_j)$ 代表节点 H_i 和 H_j 之间的距离测量值，$\hat{D}(H_i,H_j)$ 代表节点 H_i 和 H_j 之间的距离预测值，根据关系式 $\forall a,b,c,d\left(a,b,c,d>0\rightarrow\frac{c+d}{a+b}\leq\max\left(\frac{c}{a},\frac{d}{b}\right)\right)$ 可知

$$\frac{\hat{D}(H_A,D_A)+\hat{D}(D_A,D_B)}{D(H_A,D_A)+D(D_A,D_B)}\leq \max\left\{\frac{\hat{D}(H_A,D_A)}{D(H_A,D_A)},\frac{\hat{D}(D_A,D_B)}{D(D_A,D_B)}\right\} \tag{8}$$

同理，

$$\frac{\hat{D}(D_A,D_B)+\hat{D}(H_B,D_B)}{D(D_A,D_B)+D(H_B,D_B)}\leq \max\left\{\frac{\hat{D}(D_A,D_B)}{D(D_A,D_B)},\frac{\hat{D}(H_B,D_B)}{D(H_B,D_B)}\right\} \tag{9}$$

因此，由式(8)和(9)可得

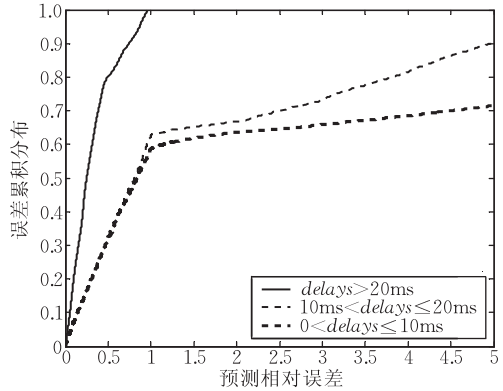
$$\frac{\hat{D}(H_A,D_A)+\hat{D}(D_A,D_B)+\hat{D}(H_A,D_A)}{D(H_A,D_A)+D(D_A,D_B)+D(H_B,D_B)}\leq \max\left\{\frac{\hat{D}(H_A,D_A)}{D(H_A,D_A)},\frac{\hat{D}(D_A,D_B)}{D(D_A,D_B)},\frac{\hat{D}(H_B,D_B)}{D(H_B,D_B)}\right\} \tag{10}$$

也就是说，在 NetPharos 中，不同预测域内预测结果迭加后的预测精度不会低于单域最差预测精度，而通过划分预测域，一定程度上避免了长距离和短距离的相互干扰，提高了单域内预测精度，从而实现最终距离预测精度的提高.

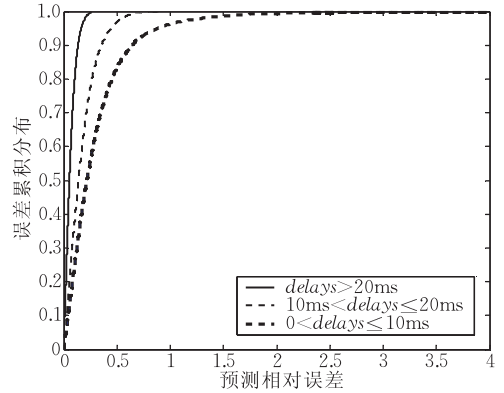
4.2 仿真分析

为了对比分析 NetPharos 与单域预测机制 GNP 之间的预测精度，我们使用 BRITE[17] 以 Waxman 随机拓扑生成算法生成一个两层次网络拓扑，其中核心网络具有 50 个节点，节点间链路带宽 10Gbps；50 个边缘网络中每个具有 100 个节点，链路带宽 100Mbps；边缘网络到核心网络的接入带宽设置为 10Mbps. 由于我们的研究目标是对比 NetPharos 与 GNP 针对静态时延的预测精度，因此采用随机分布方式为不同链路分配时延值[7,9]. 定义核心网络每条链路时延在 0.5ms 到 2.5ms 之间均匀分布，边缘网络每条链路时延在 2.5ms 到 6.5ms 之间均匀分布，边缘网络到核心网络的接入链路时延在 20ms 到 30ms 之间均匀分布. 以该网络拓扑为基础获取全部节点间的距离矩阵，分别采用 NetPharos 和 GNP 机制对其进行距离预测. 预测过程中每次构造坐标系统均随机选择 10 个节点作为基准节点，同时所有预测域均采用 5 维欧氏空间作为嵌入空间.

基于距离预测结果，通过式(5)计算两种距离预测机制的单向相对误差，图 5 分别描述了 GNP 和 NetPharos 针对不同距离范围的距离预测单向相对误差累积分布图，由图 5(a)可以看出，对 GNP 而言，短距离($delay<20$ms)的预测精度远远低于长距离($delay>20$ms)的预测精度，说明在 GNP 中，由于采用了单一度量空间嵌入，无法有效反映网络结构特征，从而导致短距离与长距离之间较大的预测精度差异. 在图 5(b)中，不同范围的距离预测精度差异不大，说明经过分层处理，NetPharos 能够很好地反应网络结构特征，避免了不同范围距离之间的相互影响. 同时，由图 5 也可以看出，不论哪个范围内的距离值，NetPharos 预测精度均优于 GNP，从而进一步验证了分层处理的有效性.



(a) GNP 预测相对误差



(b) NetPharos预测相对误差

图 5　不同距离范围预测相对误差累积分布图

为了深入分析 NetPharos 与 GNP 对不同距离范围的处理情况，图 6 描述了以 1ms 为间隔计算得到的全部距离范围内预测相对误差均值. 由该图可以看出，GNP 预测相对误差受距离值的影响很大，如相对误差最大值达到 5.1 左右，而最小值则接近 0.1. 短距离与长距离的预测相对误差之间差别非常明显，且网络距离越小，预测相对误差越大. 而 NetPharos 预测结果中则没有上述特征，无论短距离或者长距离其预测相对误差平均值全部小于 1，表明通过分层迭加机制，NetPharos 有效消除了短距离预测误差较大的问题. 考虑到网络坐标系统的一个重要作用就是帮助寻找邻居节点，提高短距离的预测精度更加有意义.

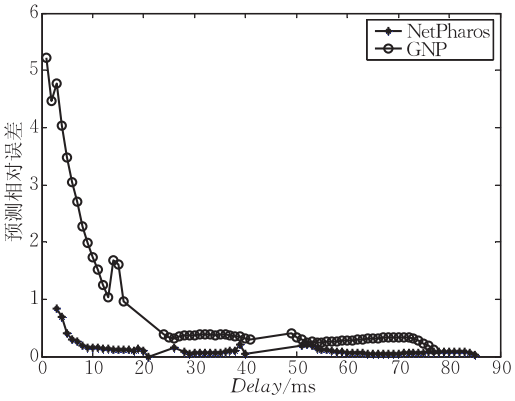


图 6 距离与预测相对误差变化图

图 7 描绘了两种距离预测机制中预测相对误差的累积分布图，其中实线代表 NetPharos，虚线代表 GNP. 由图 7 可以看出，对于 NetPharos，距离预测相对误差值小于 0.2 的节点所占比例接近 100%，而 GNP 中满足这一条件的节点只占 40%左右，可见 NetPharos 机制在相对预测误差方面明显优于 GNP. 一个可能导致这种状况的原因就是在 GNP 中，为了保证达到全局的最优，必须在全局范围内做出优化，这在一定程度上牺牲了局部最优值，而 NetPharos 内各个预测域相互独立，在确定节点坐标值时只需要满足本预测域内最优即可，同时不同域内预测结果迭加后又不会降低预测精度，因此能够使预测结果更加接近实际值. 而且根据两者的原理我们可以推断，随着网络规模的扩大，这种差别将会越来越明显.

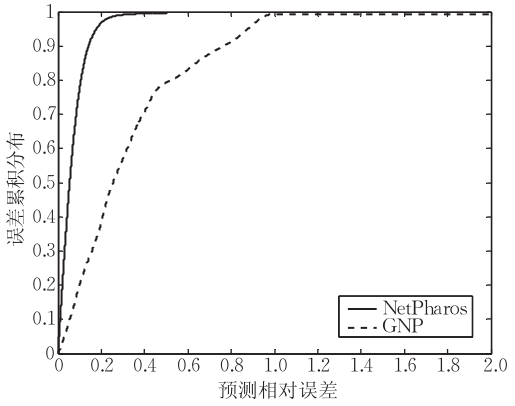


图 7 预测相对误差累计分布图

图 8 描述了两者的最近节点误判误差对比情况，根据定义，某一节点的 $CNLS$ 为 0 代表其最近节点在嵌入空间中仍然得到保留，其它数值则表示根据预测值选择的最近节点与实际最近节点之间距离相对误差值. 由图 8 可见，对于 NetPharos 而言，$CNLS$ 为 0 的节点比例约为 20%，而 GNP 中该指标数值低于 10%，说明在 NetPharos 中更多节点的最近节点在嵌入后得到了保留. 同时，对于 $CNLS>0$ 的情况，基于 NetPharos 预测结果得到的相对误差统计值同样小于 GNP，证明其预测结果优于 GNP.

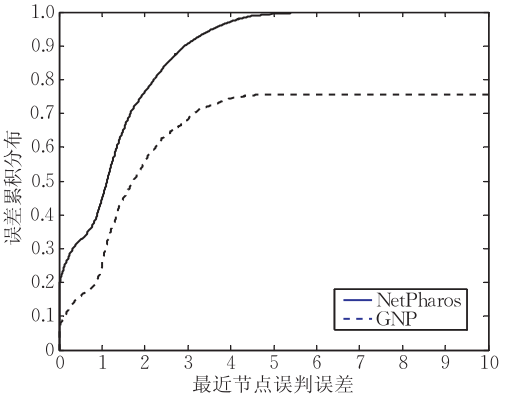


图 8 最近节点误判误差累积分布图

综合前述分析结果可以看出，在距离预测典型评价指标方面，NetPharos 性能均优于 GNP，表明 GNP 中采用单一层次的空间嵌入模式无法对网络进行精确描述，从而损失了大量可用信息，导致距离预测值无法充分反映真实网络状况. 而 NetPharos 中，由于采用了分域的层次化结构模式，能够尽可能多地保留域内信息，有效提高了距离预测的精度.

5 结束语

根据因特网层次化结构导致其性能的不均匀性，进而难以嵌入到一个均匀空间中这一状况，本文提出了一种层次化的网络距离预测机制 NetPharos. 该机制将网络划分为核心预测域和多个边缘预测域，可以对其分别采用独立的预测模型处理，既提

供了一种让 ISP 参与网络坐标系统构建的思路，又避免了预测域间的相互干扰以及传统预测方式中短距离和长距离之间的相互影响，提高了距离预测的精度.

分析与仿真结果表明，NetPharos 能够有效提高距离预测精度.但是，用于构建嵌入空间的基准节点在算法初始化过程中已经选定，这样随着时间推移，可能会出现部分基准节点失效而导致预测精度下降的现象.在下一步的工作中，需要研究将 P2P 机制引入基准节点选择过程中，设计一种基准节点随网络状况变化的动态选择机制，避免因为部分基准节点失效而导致的预测精度降低甚至失效的情况发生，进一步提高 NetPharos 的稳定性以及实用性.另一方面，基于 NetPharos 构建完善的网络距离预测系统也是我们下一步工作的重点.

参　考　文　献

[1] Stoica I, Morris R, Karger D et al. Chord: A scalable peer-to-peer lookup service for internet applications//Proceedings of the ACM SIGCOMM 2001. San Diego, 2001: 149-160

[2] Chu Y, Ganjam A, Ng T et al. Early experience with an Internet broadcast system based on overlay multicast//Proceedings of USENIX Annual Technical Conference 2004. Boston, 2004: 12-26

[3] Xu D, Kulkarni S, Rosenberg C et al. Analysis of a CDN—P2P hybrid architecture for cost-effective streaming media distribution. Multimedia Systems, 2006, 11(4): 383-399

[4] Ren S, Guo L, Zhang X. ASAP: An AS-aware peer-relay protocol for high quality VoIP//Proceedings of the 26th IEEE ICDCS, Lisboa, 2006: 70-79

[5] Ng T, Zhang H. Predicting Internet network distance with coordinates-based approaches//Proceedings of the IEEE INFOCOM 2002. New York, 2002: 170-179

[6] Shavitt Y, Tankel T. Big-Bang simulation for embedding network distances in euclidean space. IEEE/ACM Transactions on Networking, 2004, 12(6): 993-1006

[7] Costa M, Castro M, Rowstron A et al. PIC: Practical Internet coordinates for distance estimation//Proceedings of the 24th IEEE ICDCS. Tokyo, 2006: 178-187

[8] Dabek F, Cox R, Kaashoek F et al. Vivaldi: A decentralized network coordinate system//Proceedings of the ACM SIGCOMM 2004. Portland, 2004: 15-26

[9] Shavitt Y, Tankel T. Hyperbolic embedding of Internet graph for distance estimation and overlay construction. IEEE/ACM Transactions on Networking, 2008, 16(1): 25-36

[10] Lim H, Hou J, Choi C. Constructing an Internet coordinate system based on delay measurement. IEEE/ACM Transactions on Networking, 2005, 13(3): 513-525

[11] Zhang R, Hu C, Lin X et al. A hierarchical approach to Internet distance prediction//Proceedings of the 26th IEEE ICDCS. Lisboa, 2006: 73-80

[12] Li L, Alderson D, Willinger W et al. A first principles approach to understanding the Internet's router-level topology//Proceedings of the ACM SIGCOMM 2004. Portland, 2004: 3-14

[13] Dischinger M, Haeberlen A, Gummadi K et al. Characterizing residential broadband networks//Proceedings of the ACM Internet Measurement Conference 2007. San Diego, 2007: 43-56

[14] Lumezanu C, Baden R, Spring N et al. Triangle inequality and routing policy violations in the Internet//Proceedings of the Passive Active Measurement Conference 2009. Seoul, 2009: 45-54

[15] Hu N. Network monitoring and diagnosis based on available bandwidth measurement[D]. Carnegie Mellon University, Pittsburgh, 2006

[16] Elmokashfi A, Kleis M, Popescu A. NetForecast: A delay prediction scheme for provider controlled networks//Proceedings of the IEEE GLOBECOM 2007. Washington DC, 2007: 502-507

[17] Medina A, Lakhina A, Matta I et al. BRITE: An approach to universal topology generation//Proceedings of the 9th International Symposium in Modeling, Analysis and Simulation of Computer and Telecommunication Systems. Cincinnati, 2001: 346-353

XING Chang-You, born in 1982, Ph. D.. His research interests include network measurement and distributed systems, etc.

CHEN Ming, born in 1956, professor, Ph. D. supervisor. His research interests include network measurement, network management, distributed systems and novel network architecture, etc.

Background

Internet applications such as P2P, Grid and CDN can benefit much from the ability to obtain network performance before data transferring. However, Internet is designed to provide best effort packet delivery service to applications, with little or no information about the likely performance or reliability characteristics of different paths. Therefore, a suitable performance service model is needed to provide network performance for network applications based on partial measurements. Motivated by this idea, Ng firstly gave the concept of predicting network distance by coordinate embedding and proposed a Euclid-embedding based distance predication mechanism GNP. Zhang gave a detailed study on Internet distance predication error, and proposed a hierarchical distance predication mechanism. However, the hierarchy criteria used there was too simple to represent real Internet structure. There are still many other distance prediction mechanisms, such as BBS, PIC, Lighthouse, Vivaldi, and so on, but still none of them take Internet structure into consideration.

Internet forms a hierarchical structure, and different hierarchy has different network behavior mode. How to take full advantage of this hierarchical structure during distance predication process is still an open problem. The authors try to answer this question. In this paper, the authors propose a hierarchical network distance predication mechanism named NetPharos, in which many independent predication regions are constructed according to Internet hierarchical structure, and each predication region is embedded into an independent metric space. The distance between two Internet nodes can be computed by accumulation of distance values in related predication regions. By this means, the problem that the interference between short distances and long distances during distance prediction process is avoided. In addition, since predication regions are independent with each other, we can choose the most appropriate embedding coordinate space for each predication region according to its network behavior mode, which increases the system flexibility greatly.

The research is supported by the National Natural Science Foundation of China under grant No. 90304016 and National High Technology Development Program of China under grant No. 2007AA01Z418.