ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.20, No.6, June 2009, pp.1574−1590
doi: 10.3724/SP.J.1001.2009.03530

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

网络距离预测技术研究*

王意洁+, 李小勇

(国防科学技术大学 计算机学院 并行与分布处理国家重点实验室,湖南 长沙 410073)

Network Distance Prediction Technology Research

WANG Yi-Jie+, LI Xiao-Yong

(National Key Laboratory for Parallel and Distributed Processing, School of Computer, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: E-mail: sayingxmu@163.com

Wang YJ, Li XY. Network distance prediction technology research. *Journal of Software*, 2009,20(6): 1574−1590. http://www.jos.org.cn/1000-9825/3530.htm

Abstract: The distance information between nodes in P2P network is the basis for achieving topology-awareness which aims at optimizing the applications of overlay and solving the problems such as network monitoring. However, it seems infeasible to accurately and completely measure the distances between nodes due to the characteristics of P2P, such as being large-scale, self-organized, highly dynamic and so on. Consequently, researchers have put forward various prediction methods, and currently the network distance prediction technology is emerging as a new hotspot of research in P2P area. Firstly, a research framework is proposed, based on which the main aspects and the related technical issues of the research are analyzed. Meanwhile, the research history and the analysis of the classification are investigated. Many typical methods are introduced and compared. Lastly, the metrics of precision, as well as future research trends of network distance prediction is reviewed.

Key words: network distance; network coordinate; distance prediction; latency prediction; coordinate computing; topology-awareness; overlay

摘 要: P2P 网络中节点间的距离信息是实现拓扑感知以优化覆盖网应用以及解决网络监管等问题的基础.P2P 网络的大规模、自组织、高度动态等复杂特征使得要准确、完全地测量节点间的距离信息面临着极大的困难.因此,研究者们提出各种预测技术,目前对网络距离预测技术的研究已成为 P2P 领域的研究热点.首先,提出了一个网络距离预测技术的研究框架,指出了研究的重点以及相关技术问题,分析了研究历史;其次,对各种预测方法加以分类,在分类的基础上,介绍了各种典型的预测方法并进行了对比分析;最后总结了各种精确性度量标准,并指出了未来的研究趋势.

关键词: 网络距离;网络坐标;距离预测;延迟预测;坐标计算;拓扑感知;覆盖网

中图法分类号: TP393 **文献标识码**: A

* Supported by the National Natural Science Foundation of China under Grant Nos.60621003, 60873215 (国家自然科学基金); the National Basic Research Program of China under Grant No.2005CB321801 (国家重点基础研究发展计划(973)); the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant No.200141 (高等学校全国优秀博士学位论文作者专项)

Received 2008-07-02; Revised 2008-10-06; Accepted 2008-11-17

近年来,对覆盖网(overlay)的研究已经成为P2P(peer-to-peer)领域的研究热点之一.已出现许多基于覆盖网的应用,如Napster,Gnutella,OCEANSTORE[1],Kademlia[2]等文件共享系统;SplitStream[3],PROMISE[4],Bullet[5]等大型数据分发系统;基于分布式哈希表(DHT)的结构化P2P[6−10]应用以及内容分发网络(CDN)等.上述应用利用基于拓扑感知的相关技术,如覆盖网路由[11]、应用层组播[12,13]、拓扑聚合[14]、最近服务器选择[15−17]、拓扑感知覆盖网构建[17−19]等,均能显著地提高性能,因此,研究拓扑感知具有明显的现实意义.

拓扑感知的目的在于降低覆盖网的延迟伸展率(latency stretch,定义为在逻辑覆盖网中节点间的延迟与在物理网络中对应节点间的延迟之比),从而需要可扩展地获取底层物理网络节点之间的延迟信息.可见,网络距离(network distance,通常指网络延迟,一般以 round-trip time,即 RTT 表示)是实现拓扑感知的基础;同时它也是其他基于网络测量的应用,如网络监控、网络诊断等的基础.

然而,随着网络规模的不断扩大,对网络中的所有节点进行完全的距离测量在现实中面临着极大的困难:网络节点数量庞大,使得完全距离测量极易造成严重的带宽消耗和网络拥塞,同时测量的通信开销巨大;网络中的节点存在高度动态性、不可达性等问题,使得直接测量有时根本无法进行.因此,越来越多的研究者们提出采用网络距离预测技术,以减少测量开销并且使得预测精度足以满足实际应用所需.

1 网络距离预测技术研究概述

本节在提出一个网络距离预测技术研究框架的基础上,指出了网络距离预测技术研究的主要方面以及涉及的技术难题,然后分析了技术发展的历史过程,最后提出了各种基于不同划分标准的分类方法.

1.1 研究框架

网络距离预测技术研究目前尚处于研究的初级阶段,虽然受到学术界的广泛关注,然而大多数的研究还停留在理论探索阶段,更深层次的理论证明和实际部署仍是当前和未来研究工作的重点.网络距离预测技术研究需要综合考虑各方面的因素,概括其总体研究框架如图 1 所示.

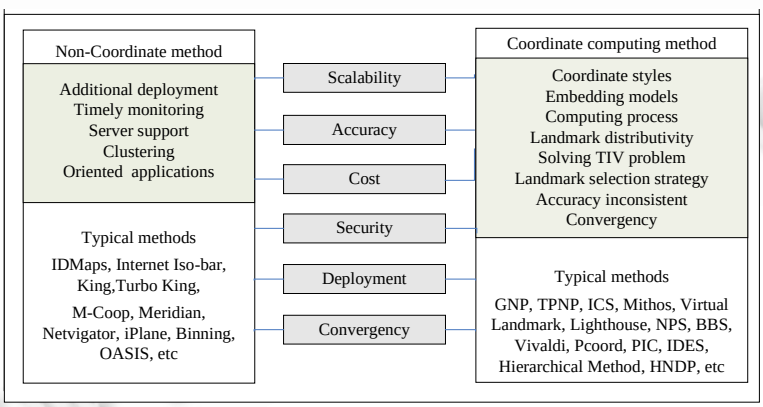

Fig.1 Framework of network distance prediction technology research

图 1 网络距离预测技术研究框架

由图 1 可知,网络距离预测技术的研究主要关注于精确性、可扩展性、测量开销(包括计算开销和通信开销)、安全性、收敛性、实际部署等方面.

根据网络距离预测技术是否采用坐标计算的形式,可以将其划分为非坐标距离预测方法和网络坐标计算方法两大类.这种分类方法从宏观角度整体上涵盖了所有的预测技术,而且能够十分鲜明地刻画两类方法的本质区别,这也是目前学术界的基本观点,因此在图 1 的研究框架中采用此分类方法.非坐标距离预测方法主要有:IDMaps[20],Internet Iso-bar[21],DDM[16],King[22],Turbo King[23],M-Coop[24],Meridian[25],Netvigator[26]等;网络坐标计算方法主要有:Triangulate Heuristic[27],GNP[28],ICS[29],Virtual Landmark[30],NPS[31],Mithos[32],Lighthouse[33],

Hierarchical Method[34],BBS(euclidean and hyperbolic)[35−37], Vivaldi[38],PIC[39],IDES[40],PCoord[41,42],HNDP[43]等.

此外,以上两类预测方法所关注的具体问题不同,非坐标预测方法关注于为实现预测所需要的额外代理节点部署、实时监控、额外服务器支持、面向的具体应用、分簇性以及分簇策略等问题;而网络坐标计算方法则关注于坐标的类型、坐标所嵌入的空间模型、坐标计算过程、三角不等性违例(triangle inequality violation,简称 TIV)、地标(landmark)的分布性、地标选择策略、精度不一致性(accuracy inconsistent)以及收敛性、安全性等方面的研究.图 2 描述了影响网络距离预测技术实用性的主要方面和相关的影响因素.

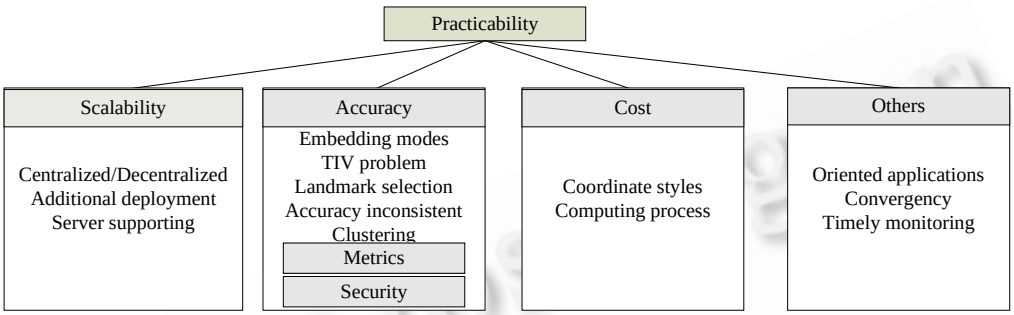

Fig.2 Main aspects that affect the practicability of network distance prediction technology and the related influencing factors

图 2 影响网络距离预测技术实用性的主要方面和相关影响因素

由图 2 可知,网络距离预测方法的综合实用性能主要取决于方法的精确性、可扩展性、测量开销以及方法的收敛性、实际部署、实时监控等方面.可扩展性方面的主要影响因素有:网络坐标计算方法中地标的分布性,包括集中式(centralized)和分布式(decentralized)地标;在非坐标距离预测方法中,额外代理节点的部署和相关服务器的支持等.精确性方面的主要因素有:坐标嵌入模型,包括欧氏嵌入(Euclidean embedding)和非欧氏嵌入(non-Euclidean embedding)两类[44];三角不等性违例TIV现象;地标选择策略,包括随机(random)选择、最近(closest)选择及混合(hybrid)选择 3 种[39];精度不一致问题,是指对于长距离和短距离节点坐标预测精度不能到达一致,从而不能达到全局最优的问题;分簇性,分簇能够有效提高网络距离预测的精度,然而不同的分簇方法对预测的结果会产生较大偏差;安全性,特别是恶意节点的干扰、破坏等;度量标准(metrics),同一种方法采用不同的度量标准来考察预测精度,结果往往不同.在测量开销方面:主要包括通信开销和计算开销两部分,对于非坐标预测方法,其计算开销均相对较小,通信开销则取决于不同的探测策略;对于坐标预测方法,通信开销取决于探测包的通信形式,如Piggy-back方式,其通信开销较其他方式较小,而计算开销主要受坐标类型以及坐标计算过程的影响.坐标类型包括绝对坐标和相对坐标[45]两种,由于两种坐标具有本质的不同,它们对应不同的坐标类型,计算开销相差较大;坐标计算过程,包括普通数学计算方法以及模拟物理过程方法,其中后者的坐标计算精度通常优于前者,而坐标计算开销一般小于前者.

1.2 研究历史

早期的研究者们受到某些网络应用,特别是邻近服务器的选择等问题的驱动,需要了解网络节点间的距离信息.然而,由前面的分析可知,实现完全的点到点之间的距离测量在 Internet 环境下是极其困难的,甚至是不可能的,因此才有了各种网络距离预测技术的诞生.表 1 按照时间顺序,列举了部分典型的网络距离预测方法,以及相关的研究机构和论文发表的会议或期刊,简要地描述了距离预测技术研究的历史.需要说明的是,有关网络距离预测技术的研究远不止表 1 中所列出的,它们只是最具代表性的部分方法.

Table 1 History of network distance prediction technology research

表 1 网络距离预测技术研究历史

Time	Method	Research institution	Proceeding/Journal
2000	DDM	Stuttgart University	INFOCOM
2001	IDMaps	University of Michigan,California	TON
2002	Internet Iso-bar	U.C. Berkeley	SIGMETRICS
	GNP	Carnegie Mellon University	INFOCOM
	King	Washington University	IMC
	Binning	U.C. Berkeley	INFOCOM
2003	M-Coop	Georgia Technology Institute	WIAPP
	BBS(Euclidean)	Tel-Aviv University, Israel	INFOCOM
	ICS	Seoul National University	IMC
	Virtual landmarks	Boston University	IMC
	Lighthouse	London, Cambridge University	IPTPS
2004	Mithos	IBM Research	SIGCOMM
	NPS	Carnegie Mellon University	USENIX
	BBS(Hyperbolic)	Tel-Aviv University, Israel	INFOCOM
	Vivaldi	MIT	SIGCOMM
	PCoord	Massachusetts Institute of Technology	NCA 04 ITNG 06
	PIC	Microsoft Research, Cambridge	ICDCS
2005	Meridian	Cornell University	SIGCOMM
2006	Netvigator	Hewlett-Packard Labs	SIGCOMM
	iPlane	Washington University	IMC
	OASIS	New York, Standford University etc.	NSDI
	IDES	Pennsylvania University	SAC
	Hierarchical Method	Purdue University,	ICDCS
2007	HNDP	Nan Jing University	IFIP
2008	Turbo King	Texas A&M University	INFOCOM

目前,国内外已有很多著名的研究机构在从事网络距离预测技术的研究,相关论文也受到网络计算与通信等计算机领域最高等级会议和期刊的青睐.自 2001 年 Francis 等人提出 IDMaps,并将其作为一项基础设施服务部署于实际应用当中,让人们看到了研究距离预测技术的现实意义和价值,使得越来越多的世界著名学府和研究机构加入到网络距离预测技术研究的行列.此外,2002 年 Carnegie Mellon 大学的 Ng 等人提出的 GNP 方法,正式宣告了使用坐标的方式预测网络距离,很快使网络坐标计算成为研究的热点.

由表 1 可知,网络距离预测方法的提出主要集中在 2002 年~2006 年,这一方面受当时研究覆盖网拓扑感知等应用热潮影响以及世界各大研究机构从事距离预测研究驱动所致;另一方面,由于 2004 年经典的分布式地标计算方法 Vivaldi 的出现,推动了研究和应用的发展.此外,各种网络距离预测方法之间并不是孤立的,往往存在某种内在的联系,研究者们总是根据相关方法的不足提出新的解决方法.

1.3 技术分类

由前面的论述可知,目前网络距离预测技术划分为两类:非坐标距离预测方法和网络坐标计算方法.

1.3.1 非坐标距离预测方法

非坐标距离预测方法通常是指通过直接测量(direct measurement)的方式测量网络中部分节点的距离,然后根据这些部分测得的距离信息来直接预测所有节点之间的距离信息,并不使用任何坐标计算的形式来预测.

此类方法通常需要在网络中部署一些特殊的节点来辅助完成预测过程,这些特殊节点可以是一些代理节点或者服务器节点.根据特殊节点的类型,可将其分为基于代理节点组合方法、基于 DNS 服务器方法以及未利用代理节点和服务器等的其他方法.此外,根据非坐标距离预测方法针对的应用目标侧重点不同,又可将其分为距离值估计与邻近性(proximity)估计两类.

- 基于代理节点组合方法、基于 DNS 服务器方法及其他方法

基于代理节点组合方法通常在现实 Internet 环境中部署一些代理节点来辅助完成所有节点间的距离预测.不同的预测方法往往采用不同的代理节点,有的采用跟踪节点(tracer),有的则采用监控节点(monitor)等.基于代理节点组合的预测方法主要有 IDMaps,Internet Iso-bar,DDM 等.

基于 DNS 服务器方法不像基于代理节点组合方法那样需要部署额外的代理节点,而是直接利用 Internet

中广泛分布的 DNS 服务器来完成距离预测.基于 DNS 服务器的主要有:King 及其改进方法 Turbo King 等.

除上述两类方法以外,有些方法既不需要部署额外节点,也不需要利用网络中的 DNS 服务器来预测节点间的网络距离.这种非坐标距离预测方法主要有:M-Coop,Meridian,Netvigator 等.

- 距离值估计与邻近性估计方法

距离值估计方法主要侧重于实时地收集网络中节点的距离信息,考察距离预测值与实际距离的误差,追求的是全局预测误差的最小化,这类方法主要有 IDMaps,Internet Iso-bar,King,Turbo King 等.而邻近性估计方法的目标不是侧重于全局估计值的精确,而是侧重于邻近节点的选择,它关心的是局部信息的正确性,此类方法主要有:Meridian,Netvigator 等.

1.3.2 网络坐标计算方法

对网络坐标计算方法的研究是网络距离预测技术研究的主体,其基本思想是通过某种网络嵌入的方式,将整个网络嵌入到一定的几何空间中,网络中的节点对应于所嵌入空间中的点,空间中节点之间的距离对应于网络中节点之间的距离,并以一定的坐标形式计算节点的网络坐标.

网络坐标计算方法很多,目前尚无严格的分类标准.根据不同的研究层面,可以采用不同的分类标准.根据预测方法中地标的分布性,可以分为集中式地标和分布式地标方法;根据所使用的坐标类型,可以分为绝对坐标方法和相对坐标方法;根据网络所嵌入的空间模型,可以分为欧氏嵌入和非欧氏嵌入方法两类[44];根据坐标的计算过程,还可以分为模拟物理过程方法和普通数学计算方法等.

- 集中式与分布式地标方法

在集中式地标方法中,地标节点通常是集中、固定的少数节点,这些节点通常是在分布性、处理能力、安全性等综合性能上较优的节点.典型的集中式地标方法有 GNP,ICS,Virtual Landmark 等.

分布式地标方法通常不需要固定的地标节点,通常任何已计算出坐标的节点均可成为被其他节点利用的地标节点.分布式地标计算方法主要有 Mithos,Lighthouse,Hierarchical Method,BBS(Euclidean and hyperbolic),Vivaldi,PIC,IDES,Pcoord,HNDP 等.

- 绝对坐标与相对坐标方法

绝对坐标是相对于相对坐标而言的,相对坐标最早由Southern California大学的Hotz[46]提出,其原理为:选择N维坐标空间中的N个节点$B_i(1\le i\le N)$作为基础节点(base nodes),则节点H的相对坐标为N维坐标,且第I维为其到第 $i(1\le i\le N)$个基础节点的距离值,表示为 $(d_{HB_1},d_{HB_2},...,d_{HB_N})$.相对坐标是相对于 N 个基础节点而言的,故称为相对坐标,经典的 Lipschitz 嵌入便利用了相对坐标的思想.相对坐标方法主要有:Triangulate Heuristic,ICS,Virtual Landmark,IDES 等,而绝对坐标方法则包括 GNP,NPS,Vivaldi 等相对坐标以外的所有网络坐标计算方法.

- 欧氏嵌入与非欧氏嵌入方法

目前大多数的网络坐标计算方法均采用欧氏嵌入方法,然而,为了解决欧氏空间中存在的相关局限性问题,不少研究者提出采用非欧氏空间嵌入方法,典型的如BBS(hyperbolic)[37]和带有高度向量的Vivaldi[38]方法等,研究表明,非欧氏嵌入方法在一定程度上能够提高预测的精度[44].

- 模拟物理过程方法与普通数学计算方法

模拟物理过程方法是指在网络坐标计算的过程中依据了相关的物理原理,该类方法典型的有 BBS,Vivaldi 和 PCoord. Vivaldi 和 BBS 分别通过模拟物理中的弹簧力场和粒子力场的方式,而 PCoord 采用摩擦力机制来调整坐标位置.然而,网络坐标计算方法的大多数属于普通数学计算方法,其主要计算过程是依靠地标并借助相关的数学知识计算坐标.

综上所述,可将网络距离预测技术的整个分类方法如图 3 所示来描述.

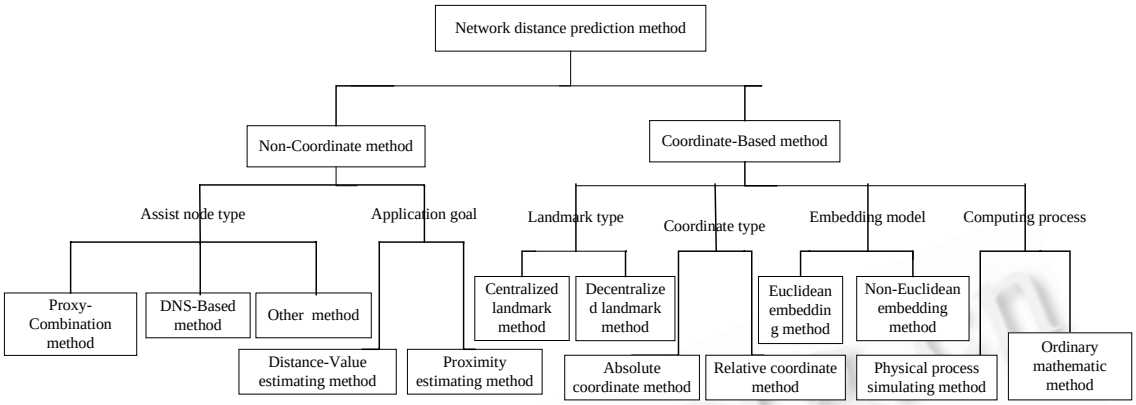

Fig.3 Classification of network distance prediction technology

图 3 网络距离预测技术分类

2 非坐标距离预测方法

非坐标距离预测方法作为距离预测研究的重要组成部分,在实际应用中发挥着重要作用.这类方法通常借助于一定的辅助节点或者服务器来实时地探测部分节点间的距离,从而依据一定的算法估计所有节点间的距离.根据应用目标侧重点的不同,可以分为距离值估计和邻近性估计两种.这种分类方法以目标为主要关注点,已得到多数研究者的认同.下面以此分类为基础,就目前最典型的非坐标距离预测技术加以介绍和分析.

2.1 距离值估计方法

距离值估计方法主要关注节点间延迟的值估计.这种值估计的过程是基于部分已知局部延迟信息的,而这种局部距离信息的测量往往需要一定的代理节点或者服务器作支撑,例如 IDMaps,Internet Iso-bar,DDM 等方法使用了额外部署的代理节点.

IDMaps 方法通过在不同的自治系统(autonomous system,简称 AS)附近部署一些跟踪节点(tracers),从而形成各 Tracers 之间以及 AS 与 Tracers 之间的双层结构,其拓扑结构如图 4 所示.双层结构中的节点互相探测距离,并将探测的结果存入称为 HOPS 的服务器中.主机之间的距离估计采用了三角启发式原理:主机 A,B 之间的预测距离为 A 到最近的 Tracer T1 的距离、B 到其最近 Tracer T2 的距离以及 T1 到 T2 之间最短路径距离的总和.

Internet Iso-bar是 2002 年UC Berkeley分校的Chen等人提出的一种基于分簇的覆盖网距离预测方法,其拓扑结构如图 5 所示.它首先采用相似性分簇方法将网络中的N个节点分成K个簇;其次,在每个簇的中间选择一个监控节点(monitor)来监视与簇内各节点的距离,周期性地重复探测距离并加以更新;最后,通过如下方法估测距离:若节点i,j在同一个簇内,且m为簇的监控节点,$d_{i,j}$表示节点i和节点j的距离,则$d_{i,j}=(d_{i,m}+d_{j,m})$;否则,假定节点$i,j$的监控节点分别为$m_i,m_j$,则$d_{i,j}=d_{m_i,m_j}$.

此外,基于相关代理节点预测网络距离的还有 DDM,M-Coop 等.DDM 与 IDMaps 相似,不同之处在于 DDM 将 Tracers 组织成层次结构,用户节点由上至下遍历各个层次来定位最近的 Tracer;M-Coop 则通过抽取 BGP(border gateway protocol)报告中的信息来获得网络节点的 AS 拓扑图,每一个节点测量到一组小数目的其他节点的距离,当两个 IP 地址之间的距离需要估计的时候,递归地进行多个测量来获得相应的估计值.

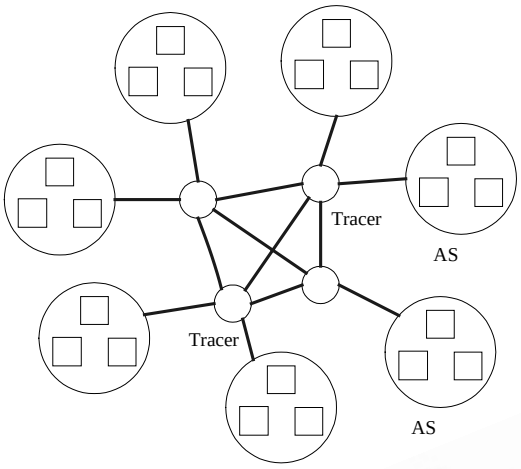

图 4 IDMaps 拓扑结构
Fig.4 Topology of IDMaps

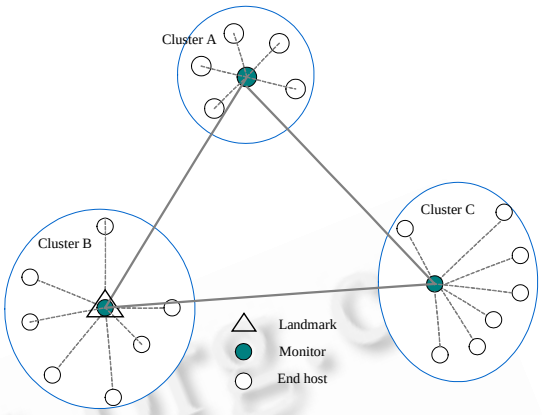

图 5 Internet Iso-bar 拓扑结构
Fig.5 Topology of Internet Iso-bar

综上可知,各种基于代理节点方法的特点如下:第一,方法一般采取相关的分簇方式,如 IDMaps 和 DDM 通过 IP 地址前缀分簇、M-Coop 通过 BGP 报告信息分簇、Internet Iso-bar 通过距离的相似性分簇;第二,通过代理节点维护相关的拓扑结构,如 IDMaps 和 DDM 通过 Tracer 维护局部 AS 图、利用 HOPS 服务器维护全局 Internet 虚拟拓扑图,Internet Iso-bar 通过 Monitor 维护局部拓扑图;第三,适当增加代理节点的数目,能够提高距离预测的精度,但同时会增加维护和管理的开销;第四,此类方法一般采取局部集中、全局对等的方式来预测距离,容易导致单点失效、无法预测等问题.

King 方法与上述方法不同,它不需要额外部署任何代理节点或者监控节点,而是利用 Internet 中广泛分布的 DNS 服务器,是目前使用最多的一种在线的测量方式.King 通过递归地查询 DNS 方式实时地估算节点间延迟,估算的方法是将与两个节点距离最近的 DNS 之间的延迟当作这两个节点之间的距离,如图 6 所示.

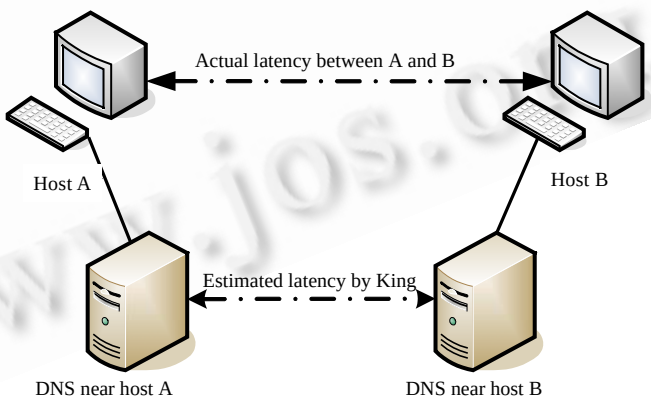

Fig.6 Measuring principle of King method
图 6 King 方法的测量原理

2008 年,Leonard 等人在详细分析 King 的精确性、可扩展性不高等不足的基础上提出了改进版 Turbo King,它主要解决了 King 中存在的因需要在远程服务器中插入无数冗余的 DNS 记录信息和在大规模部署当中因为巨大的通信开销所导致的缓存污染和扩展性不高等问题.

综合上述距离值估计方法,其共同点是:第一,方法多为早期提出的基于直接测量方式的预测方法,实时性

较好,但其距离值是基于某种间接方式获得的,精确性相对不高;第二,需要了解局部或者全局拓扑信息,因此在测量的过程中,往往需要知道源节点和目标节点的 IP 地址信息,这使得在目标节点 IP 不知道的情况下无法预测;第三,此类方法需要一定的辅助节点或者服务器的支持,由于此类特殊节点的存在,易造成单点瓶颈现象,影响方法的可扩展性;第四,此类方法由于需要实时收集更新测量节点间的探测结果,存储开销和通信开销较大.

此外,需要强调的是 King 方法的特殊性:一方面,IDMaps,Internet Iso-bar,DDM 都是以系统的方式提出来的,而 King 方法则是以一种技术的形式提出来的,King 方法关注的是在 Internet 中如何有效地得到节点间的距离,因此该方法的实用性很强,可以作为一种技术手段运用于其他预测方法当中;另一方面,由于 King 方法不需要额外部署相关节点,也不需要集中式服务器支持,故其可扩展性相对较好.此外,King 方法是一种在线的直接测量方法,不需要通过相关的外推计算获得距离值.

2.2 邻近性估计方法

此类非坐标距离预测方法从实际应用的角度出发,重点关注于从多个参考节点中选出邻近的节点,而非测量出具体的距离值,典型的有Meridian,Netvigator,Binning,iPlane,OASIS等方法.这些方法的技术着眼点在于如何利用新技术优化最优节点的查询,而非距离值的最优测量;在最优查询的过程中,可能借助一定的几何或者空间结构来组织节点以达到查询的目的,如Meridian的同心环、Binning[17]中的分箱结构;也可能借助一定的辅助信息来指导查询,如Netvigator中的中间路由器探测信息、iPlane[46,47]与S3[26]中的网络拓扑信息、OASIS[48]中的地理位置信息等.

Meridian 由 Carnegie Mellon 大学的 Wong 等人于 2005 年提出,它利用基于直接测量形成的松散结构的同心环状结构,结合 Gossip 协议来交换节点间的信息,以迭代转发查询匹配信息的方式来定位最优的邻近节点.2002 年,Ratnasamy 等人提出使用装箱(binning)方式进行邻近性预测,并用于覆盖网构建等.它首先通过探测至部分地标节点的距离,获得相应的距离向量,然后将具有相似距离向量的节点划分至相同的箱中以达到分簇的目的.

惠普实验室的 Sharma 等人在 2006 年提出了 Netvigator 方法,通过探测小数目的地标节点和中间路由器(称作 Milestones)来检测最近的节点.该方法的创新之处在于通过增加探测节点路径上的 Milestones 信息来避免如 Binning 等人方法中的错误分簇(false clustering,源于具有相同地标向量的节点为邻近节点所致)问题,因此提高了局部网络特征信息的精度.

此外,Chen等人提出的WNMS[49]方法研究通过监控节点的最优放置来优化邻近性预测;iPlane[46,47]和S3[26]提出使用根据Traceroute测量收集的网络拓扑信息来辅助邻近选择;OASIS[48]系统使用地理位置信息进行邻近选择等.

以上邻近性估计方法针对实际应用,从不同的技术角度研究了如何遴选最近节点的问题,方法的目标明确;与距离值估计方法相比,不仅降低了存储开销和通信开销,而且方法的实际部署开销更小,预测精度更高.

综合上述所有典型的非坐标距离预测方法,从不同角度对它们进行对比,结果见表 2,其中 N 为主机(hosts)数,C 为簇的数目,K 为地标数,M 为 Meridian 方法中环的数目.

3 网络坐标计算方法

网络坐标计算方法目前已成为距离预测技术研究的主流方法.它通过将网络中的节点映射到一定的几何空间当中,通过坐标来定位网络中的节点在几何空间中的位置,并且根据节点的坐标信息来计算节点间的距离.

最早使用坐标来预测距离的方法是基于相对坐标的三角启发式(triangulate heuristic)方法,可描述为:假定系统中存在$H1$,$H2$两个节点,其相对坐标分别为 $(d_{H_1B_1},d_{H_1B_2},...,d_{H_1B_N})$ 和 $(d_{H_2B_1},d_{H_2B_2},...,d_{H_2B_N})$,其中$d_{ij}$表示节点$i$,$j$的探测距离,则它们距离的下限为 $L=\mathrm{Max}_{i\in(1,2,...,N)}\left(\left|d_{H_1B_i}-d_{H_2B_i}\right|\right)$,上限为 $U=\mathrm{Min}_{i\in(1,2,...,N)}\left(\left|d_{H_1B_i}+d_{H_2B_i}\right|\right)$.根据$L$,$U$或者$L$,$U$的加权便可估计节点间的距离,如Guyton和Schwartz[27]曾使用(L+U)/2 估计距离,并以此解决最近服务器选择问题,然而结果表明其效果并不理想,它只是一种最原始的以坐标形式预测距离的方法.

Table 2 Comparison of different methods that based on direct measurement

表 2 不同的非坐标预测方法之间的对比

Property	IDMaps	Internet Iso-bar	King Turbo King	M-Coop	Meridian	Netvigator
Main technique	Triangulation inequality, proximity-based clustering	Similarity-Based clustering	Using nearby DNSs distance	Active measure with AS graph and BGP reports	Multi-Resolution rings overlay structure and gossip protocols	Using milestones to accurately locating closest nodes
Measure cost	$O(C^2+N)$	$O(C^2+N)$	$O(N^2)$	$O(C^2)$	$O(M*\log N)$	$K*N$
Scalability	Restricted	Restricted	High	High	High	High
Accuracy	Low-accurate	Medium	Medium	Medium	High	High
Timely	Yes	Yes	Yes	Yes	No	No
Additional deployment	Tracers and HOPS	Monitor	DNS (Internet exists)	Transit AS's	No	Milestones (Internet exists)
Clustering	Yes	Yes	No	No	Yes	Yes
Primary application examples	Nearest mirror selection	Overlay location and routing	Constructing topologically sensitive overlay	Improve content distribution and overlay multicast	Closest node discovery, locating nodes with latency constraints	Proximity selection

真正意义上开始利用坐标计算方式预测网络距离的是 Ng 等人提出的 GNP 方法.自从该方法提出以后,网络坐标计算正式成为距离预测技术研究的热点.

3.1 集中式地标方法

在集中式地标方法中,地标通常是集中固定的预先选取好的少数节点,节点的性能相对较高,并且地标节点的分布性一般较好.此类方法有其固有的优势:一方面,新加入的节点可以直接访问已部署的地标节点,从而减少了新加入节点查询地标的开销;另一方面,由于固定地标的性能以及分布性较好,坐标计算的精度较高.然而集中式的方法也具有致命的缺点:集中式的地标节点易成为通信、负载瓶颈,造成单点失效问题,扩展性明显受到限制,而且易成为恶意攻击的对象,安全性难以得到保证.典型的集中式地标方法有 GNP,ICS,Virtual Landmarks 等.

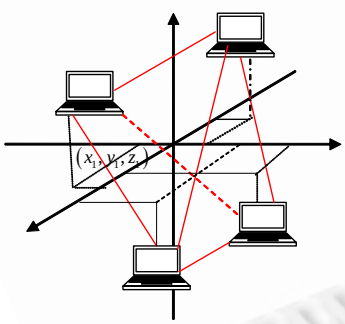

Fig.7 Geometric model of Internet space

图 7 Internet 的几何空间模型

GNP将Internet建模为一定维度的几何空间(如图 7 所示),将网络中的主机节点映射为欧氏空间中的点,通过绝对坐标计算方法来预测网络距离.其坐标计算分为两步:首先计算地标节点的坐标,然后依据地标节点的坐标和至地标的探测距离计算普通节点的坐标.坐标计算的实质为求解估计误差与实际测量误差的一般多维全局最小化问题,目前有很多方法可以进行近似求解,GNP中利用Simplex Downhill方法[50]计算其坐标.GNP方法采用网络坐标的方式预测距离,与非坐标预测方法相比,大幅度减少了探测的开销以及实际部署开销,预测精度也较高.然而,它也面临着诸如集中式地标带来的单点失效以及非线性计算开销较大等问题.

与 GNP 不同,2003 年,Seoul 大学的 Lim 等人提出的 ICS 和 Boston 大学 Tang 等人提出的虚拟地标(virtual landmarks)方法采用相对坐标方法计算网络坐标.在 ICS 和 Virtual Landmarks 中,普通节点首先获得地标节点的距离矩阵信息并以距离向量形式表示,向量的维度等于地标节点的数目;其次,在基于 Lipschiz 嵌入的基础上采用 PCA(principle component analysis)技术来降低维度以提高计算效率.

Lipschiz嵌入是欧式嵌入的一种特殊类型,对于任意节点i的坐标向量$\vec{x}_i \in R^n$,则$\vec{x}_i$的第j个元素为i到地标j的距离.Lipschiz 嵌入的本质是以距离作为坐标的元素,其结果的精确性源于在度量空间中的两个相近的节点具有到其他对象的相似距离[29].

由于在距离矩阵中,多个距离向量之间可能是相似的,那么采用 PCA 技术消除相似的向量,将原向量空间重

新映射为维度更小的正交笛卡尔空间,以减少维度、简化坐标计算.由于同一簇中的节点到地标节点具有相同的距离向量,因此以上方法为网络距离预测中的分簇方法提供了理论依据,其实质是通过向量形式, 采用线性变换的方式来计算节点的坐标,其计算效率远远高于 GNP 等方法采用的非线性迭代方法求解.此外,由于降低了坐标维度,使得用于坐标计算的地标数目要求降低,因此即使在部分地标失效的前提下仍然能够进行坐标计算,因此大大增强了方法的鲁棒性和有效性.

集中式地标方法多为网络坐标计算研究早期提出的方法,此方法与非坐标距离预测方法相比效率更高,精确性更好.由于集中式地标的缘故,使得方法的可扩展性和安全性较差,因此后来的研究大多是针对分布式地标方式.以上采用绝对坐标和相对坐标的计算技术为后续研究奠定了坚实的基础.

3.2 分布式地标方法

分布式地标方法力图避免集中式地标的缺陷,通常,系统中已计算出坐标的任意节点均可作为地标,而非使用固定的节点作为地标.目前此类方法很多,代表性的有 Lighthouse,NPS,PIC,IDES,Hierarchical Method,HNDP,BBS,Vivaldi,PCoord 等.

2003 年,London 大学的 Pias 等人提出的 Lighthouse 方法采用坐标变换的方式计算网络坐标.系统中首先设定一个全局坐标基,新加入的节点首先通过随机探测 K+1 个(K 为坐标的维度)地标获得相应距离信息,并通过 Gram-Schmidt 过程计算其局部坐标基,然后通过局部与全局坐标基的转换矩阵来计算新节点的全局坐标.Lighthouse 与 ICS,Virtual Landmarks 坐标计算的方式不同,但其计算的实质相同,都是通过坐标的线性变换实现,因此其计算开销较小.然而,该方法的预测精度明显受转换矩阵以及坐标维度的影响,合适的转换矩阵以及维度的选取依然是一个严峻的问题.

NPS 是 GNP 的改进方法,与 GNP 不同的是:NPS 中所有的节点都可以作为地标,并且地标的选择是随机的,坐标的计算由节点自身计算而并非由集中的节点计算,增强了系统的可扩展性;NPS 将地标节点组织成层次结构以维护一致性问题;通过周期性地探测节点间的延迟以检测地标的稳定性;通过对恶意节点的检测以提高系统的安全性等.NPS 较早研究了坐标系统的实际构建问题,然而它只是坐标系统实际部署的初步尝试,仍然会面临许多亟待解决的问题.

2004 年微软研究院的Costa等人提出的PIC方法首先研究了地标选择策略问题,通过对比随机、最近、混合 3 种地标选择策略,并且最先发现了远距离和近距离节点的坐标计算不能达到同时精确的精度不一致性问题.同时指出混合地标策略的综合性能最好,然而Zhang等人[34]指出这种混合策略同样不能精确预测所有类型的距离.

与PIC方法类似的有,Florida大学的Onbilgar等人提出的TPNP[51]和IBM研究院的Waldvogel等人提出的Mithos[32]方法.TPNP方法采用一种纯P2P结构来消除集中式方法中的固定地标,新加入的节点首先发现已在系统中的邻近节点并将其作为地标;Mithos方法选择距离最近的部分节点作为地标进行坐标计算,该方法由于需要大量的探测以确定邻近性,因此通信开销较大.

此外,针对 PIC 提出的精度不一致性问题,2006 年 Purdue 大学的 Zhang 等人以及 2007 年南京大学的 Xing 等人分别提出了 Hierarchical Method 和 HNDP 方法.

Hierarchical Method 采用一种层次式方法利用多个地标集合计算多个坐标集,每一个坐标和地标集合对应不同的距离范围.例如,一个节点使用一个坐标来估计到远节点的距离,而用另一个坐标来估计到近节点的距离.如若发现一组坐标对近距离预测精确,而另一组对远距离是精确的,则根据感兴趣的范围,综合选择合适的坐标集合以最高的整体预测精度.HNDP 提出另一种层次式的预测方法,它将 Internet 划分成多个独立的区域(如 Edge,Core,Region,Dual 等),在各个区域内分别采用最近地标选择策略计算坐标,最后节点间的预测距离通过累加不同区域内的距离得出.

以上两种方法是基于 PIC 中得出的结论,即近距离节点的坐标预测精确,而远距离节点的预测距离不精确而提出来的.层次式方法能在一定程度上克服精度不一致性问题,然而,Hierarchical Method 中坐标集的选择、多个坐标的融合以及网络坐标计算中面临的其他问题仍然需要继续深究;而在 HNDP 方法中对实际网络进行层

次性区域划分是难以实现的,而且要知道整个 Internet 的拓扑结构也是不现实的.

网络坐标具有欧氏嵌入(Euclidean embedding)和非欧氏嵌入(non-Euclidean embedding)两种嵌入模型[44].在欧氏空间中,距离函数必须满足自反性、对称性、三角不等性的约束.目前,大多数坐标计算方法都是采用欧氏嵌入,然而,Zheng等人[52]证明在网络中存在不少TIV现象,现实Internet网络呈现出部分非欧氏性,分析并解决网络中由于TIV造成的预测不精确问题已经成为许多研究解决的热点问题,目前已有不少预测技术提出了TIV问题的解决方法[39,52−55].

最典型的是 2006 年 Pennsylvania 大学的 Yun 等人提出的 IDES 方法.该方法通过采用 SVD(singular value decomposition)和 NMF(non-negative matrix factorization)两种矩阵分解技术来建模子优化和非对称路由策略;为每个节点赋予一个入向量和出向量,根据出向量和入向量的内积确定网络距离.该方法摆脱了对称性和三角不等性的约束,它假设距离矩阵中存在大量线性相关的向量,即利用分簇性原理,当出现错误分簇时,IDES 的预测精度会明显降低.

综合以上分布式地标方法可知,在避免使用集中式地标以提高可扩展性的同时,着重围绕地标选择策略、精度不一致性、恶意节点的检测、三角不等性违例(TIV)等问题来解决.以下 3 种方法则是从坐标计算的收敛性、精确性、预测开销等方面研究网络坐标的计算问题,它们的共同点在于通过模拟相关的物理过程来优化坐标计算;将模拟特定物理过程的方法用于网络计算的思想,最早是由澳大利亚Newcastle大学的Eades[56]等人在研究图形绘制应用的弹簧嵌入算法时提出来的.

受此启发,不少研究人员也提出了一些新的模拟物理过程的坐标计算方法,如 BBS,Vivaldi,PCoord 等.

BBS(big-bang simulation)[35−37]方法模拟了粒子在由于嵌入误差所产生的力场中的爆炸行为.它将网络中的所有节点视作一个粒子集合,粒子之间可能相互吸引或者排斥,可能由于力场的作用加速,也可能由于摩擦力的作用而减速.Vivaldi方法则模拟了节点在嵌入误差产生的弹簧力场作用下的运动过程,它假定任意两个节点之间都存在一根弹簧,通过节点间弹力的作用来修正节点的坐标.PCoord方法则在坐标更新的过程中引入摩擦力机制来提高坐标的收敛速度,从而提高坐标的精确性和可用性.

BBS方法的优势在于坐标计算的精确性,它与GNP等绝对坐标方法中使用的传统的梯度最小化方法,如Downhill Simplex,Steepest Decent方法相比,能够有效解决计算迭代过程中的局部最小化(local minima)问题;在诸如子网路由、最近镜像选择、拓扑聚合等应用中效率明显,然而BBS系统相对复杂,并且需要系统的全局知识.Vivaldi则是目前公认的最有效的坐标计算方式,它采用启发式的坐标调整过程,消除了地标,以Piggy-back方式通信、非线性坐标计算等特点使得Vivaldi在精确性、可扩展性、通信和计算开销等方面优势明显;然而Vivaldi方法收敛速度较慢而且不能保证能够收敛到稳定状态,目前不少研究[57−59]指出,Vivaldi系统在收敛性和稳定性上存在问题,使得在实际应用时受到很大限制.PCoord方法采用类似GNP的方法,不过在坐标计算过程中增加了周期性的抽样和坐标更新阶段,以摩擦力机制在一定程度上提高了坐标计算的收敛速度.

一般地,模拟物理过程的分布式地标方法通过增加网络坐标的更新调整过程,主动适应网络动态变化环境,坐标的适应性、鲁棒性、精确度通常较高,然而此类方法的问题在于收敛性与稳定性问题,以及为优化收敛性、稳定性所带来的计算开销和通信开销等问题,在性能与代价方面往往需要折衷考虑.

综合以上各种具有代表性的网络坐标计算方法,根据方法提出的时间顺序,从方法对应的主要技术、地标分布性、分簇性、计算开销、TIV 问题的解决以及安全性检测方面归纳为表 3.

4 精确性度量标准

目前衡量网络距离预测精度的度量标准不一,综合起来大致有 5 种度量标准:张力(stress)、相对误差(relative error,包括绝对相对误差(absolute relative error)和有向相对误差(directional relative error))、范围精确度(range accuracy)[34]、相对次序缺失(relative rank loss,简称RRL)[44]、最近邻居缺失(closest neighbors loss,简称CNL)[17,44,60].

Table 3 Comparison of different methods that based on coordinates computing

表 3 不同的网络坐标计算方法之间的对比

Properties	Main technique	Centralized	Embedding type	Cluster	Physical process	Compute cost	Solving TIV	Security
GNP	Absolute coordinate Simplex downhill	Yes	Euclidean	No	No	Big	No	Ignore
ICS	PCA and Lipschiz embedding	Yes	Lipschiz	No	No	Small	No	Ignore
Mithos	Improve GNP with closest nodes as landmark	No	Euclidean	No	No	Big	No	Ignore
Virtual Landmark	PCA and Lipschiz	Yes	Lipschiz	No	No	Small	No	Ignore
Lighthouse	Gram-Schmidt,Vector base translation	Yes	Euclidean	No	No	Small	No	Ignore
NPS	Random index nodes selection, Landmark delaminate	No	Euclidean	No	No	Big	No	check
BBS	Simulate particle explosion	No	Euclidean/ Hyperbolic	No	Yes	Medium	Yes	Ignore
Vivaldi	Simulate spring force field and piggy-back communication	No	Euclidean /with high vector	No	Yes	Small	Yes	Ignore
PIC	Improve GNP with different landmark selection strategies	No	Euclidean	Yes	No	Small	No	check
PCoord	Active exchange information; sample weight, force of friction, Damping mechanism	No	Euclidean	Yes	Yes	Big	No	Ignore
IDES	Matrix Factorization: SVD and NMF	No	Euclidean	No	No	Small	Yes	Ignore
Hierarchical Method	Multi-Coordinates	No	Euclidean	No	No	Big	No	Ignore
HNDP	Divide Internet into many different regions	No	Euclidean	Yes	No	Big	No	Ignore

(1) Stress

Stress是用于坐标计算方法中描述坐标嵌入适应性的标准方法,若$d_{x,y}$表示x,y间的实际测量距离,$\hat{d}_{x,y}$表示预测距离,则应力可以采用如下Stress-1[61]形式表示:

$$Stress-1=\sigma 1=\sqrt{\frac{\sum_{x,y}(d_{x,y}-\hat{d}_{x,y})^2}{\sum_{x,y}d_{x,y}{}^2}}.$$

该精确性度量标准旨在描述由于坐标嵌入失真(distortion)所导致的精确度低的问题.

(2) Relative Error

Relative Error是目前绝大多数网络距离预测方法采用的精确性度量标准,若$d_{predict}$表示预测的距离,$d_{measure}$表示实际测量的距离,则绝对相对误差、有向相对误差可分别表示如下:

$$R_{abs\_error}=\frac{\left|d_{predict}-d_{measure}\right|}{\min(d_{predict},d_{measure})},\quad R_{dir\_error}=\frac{d_{predict}-d_{measure}}{\min(d_{predict},d_{measure})}.$$

相对误差描述的是预测距离与实际测量距离差值所占的比重.这种方法从整体上描述了预测值与实际值的差异大小,在对于某些具体应用时并不能给出任何指示性的帮助.

(3) Range Accuracy

Range Accuracy 描述的是在某一给定的距离范围内预测的精度问题.对于一个给定的距离 r,假定通过预测算法预测的所有距离在 r 内的路径集合为 *PredictedLinks*,而假定实际测量所得的距离在 r 内的路径集合为 *MeasuredLinks*,且对于集合 X,$|X|$表示集合容量,则 Range Accuracy 可定义为

$$RA=\frac{|MeasuredLinks\cap PredictedLinks|}{|MeasuredLinks|}.$$

类似地,可定义逆向范围精确度(reverse range accuracy,简称 RRA)为

$$RRA=\frac{|MeasuredLinks\cap PredictedLinks|}{|PredictedLinks|}.$$

RRA 反映了由于预测误差所导致的在不同距离范围内的干扰问题.*RA* 与 *RRA* 对于需要查询在一定距离范围内的所有节点的应用是理想的度量标准.

(4) Relative Rank Loss (RRL)

RRL 描述的是距离某个节点远近的次序关系维护程度,需要解决诸如:"哪个节点距离 *A* 更近,*B* 还是 *C*?"的问题.RRL 值的计算通常采用线性代数中的逆序数计算方法.假定 R_i 表示第 *i* 个数的逆序数,共有 *n* 个数,则 *RRL* 可定义为

$$RRL=\frac{\sum_1^n R_i}{C_n^2}.$$

RRL 值在 0~1 之间,0 表示节点的远近次序关系完全正确,1 表示全部错误,而 0.5 表示只有一半估计正确.

(5) Closest Neighbors Loss (CNL)

*CNL*值描述的是距离某个节点最近的节点关系保持状况.1 表示预测的、最近的节点与实际的、最近的节点相同,0 表示不同.若用*CNL*值描述系统中所有节点的最近节点维持关系,则其所描述的是整个系统中所有节点的最近节点的预测精确度,如值 0.5 表明系统中只有一半的节点其最近邻居预测正确.假定系统中共有*n*个节点,C_i 表示第*i*个节点的CNL值,当预测正确时,C_i=1,否则C_i=0,系统整体*CNL*值可表示为

$$CNL=\frac{\sum_1^n C_i}{n}.$$

综合上述各种度量标准,可将各种方法按照其计算公式、描述的意义以及目的归纳,见表 4.

Table 4 Comparison of different methods of precision metrics

表 4 不同的精确度度量方法对比

Metrics	Formula	Description	Purpose
Stress	$\sqrt{\frac{\sum_{x,y}(d_{x,y}-\hat{d}_{x,y})^2}{\sum_{x,y}d_{x,y}^2}}$	Describe the embedding distortion	To find whether the embedding is suited or not
Relative Error	$R_{abs\_error}=\frac{\lvert d_{predict}-d_{measure}\rvert}{\min(d_{predict},d_{measure})}$	Describe the discrepancy between practical and predicted distances	To find the predicted results are whether close to practical ones
RA/RAA	$\frac{\lvert MeasuredLinks\cap PredictedLinks\rvert}{\lvert PredictedLinks\rvert}$	Describe the accuracy with in some distance range	To find the accuracy of predicted paths within the range
RRL	$\frac{\sum_1^n R_i}{C_n^2}$	Describe the relative rank loss for some node	To find who is closer to me?
CNL	$CNL=\frac{\sum_1^n C_i}{n}$	Describe the accuracy of closest nodes predicted in the system	To find who is the closest to me?

5 未来研究趋势

综合网络距离预测技术研究的热点问题以及广泛关注的应用领域,认为未来研究趋势主要是在以下几方面:

(1) 网络嵌入模型的研究

对Internet建模是一项复杂的任务,因为现实Internet环境中存在诸多如高度动态性[62]、网络节点不可达性[63]以及TIV等特征.目前虽然已有欧式空间、双曲空间以及Lee等人[55]提出的混合空间等嵌入模型用于网络距

离预测的建模计算,但是这些模型都不能很精确地描述现实Internet的特征.因此,研究如何对Internet环境建模,寻找一种最优的空间模型是数学界和计算机界共同需要解决的问题.

(2) 面向应用的精确性度量标准研究

对于不同的应用,同一种预测技术采用不同的度量标准可能会得出截然相反的结果[44].目前已有不少研究人员研究了精确性度量标准问题,并且针对相关的应用提出了新的度量标准,如RRL的扩展形式SRRL、CNL的扩展形式ECNL等,基于这种度量标准所得出的结果运用于具体的应用当中,其效果明显优于传统的相对误差等度量标准得出的结果.可见,针对目前网络距离信息的不同应用,研究出能够准确刻画面向该应用的精确性度量标准也是未来研究的一个重要方面.

(3) 高效、稳定的预测方法研究

由前面的分析可知,要实现高效、稳定的网络距离预测,需要考察技术的可扩展性、计算开销、通信开销、精确性、稳定性、收敛性、安全性等诸多方面,而与之密切相关的研究主要包括:坐标类型选择、计算过程模拟、分簇算法、现实部署性等等.在网络坐标计算方法中,采用非线性方法进行绝对坐标计算其收敛性好,但是开销大;而采用线性方法进行相对坐标计算其计算开销小,却需要节点的分簇性假设;采用模拟物理过程的方法如 Vivaldi,其计算和通信开销小,但是收敛速度慢且不能保证能够收敛到稳定状态,不便于实际应用.在非坐标距离的预测方法中,代理节点或者服务器节点部署的数目、方式以及开销等现实部署性问题,还有寻找最优的分簇算法仍然值得深究.

今后的研究主要关注于:非坐标距离预测方法中的 King 方法和邻近性估计方法的研究以及网络坐标计算方法中的分布式地标方法的研究.使用基于代理节点方法预测网络距离值的方法,其实际部署开销过大,预测精度相对较低;而集中式地标方法面临着可扩展性和安全性低等问题,使其不适合当前大规模分布式的应用.此外,将分布式地标方法与非坐标距离预测方法 King 结合起来,能够提高探测的效率;将模拟物理过程方法和绝对坐标方法结合起来,能够极大地提高坐标收敛性和稳定性;将相对坐标方法与合适的地标选择策略、TIV 解决方法等融合,能够有效地改善坐标的计算开销和预测精度.

(4) 安全性方面的研究

安全性方面的研究主要针对于网络坐标计算方法,在非坐标距离预测方法中,其安全性来源于代理节点或者服务器等的安全性问题.当前许多对网络坐标安全性的研究[39,59,60,64,65]表明,网络坐标系统在面对实际Internet时是很脆弱的,极易遭受来自各方面的攻击.目前针对网络坐标安全性的研究很少,仅有少数简易的检测恶意节点的方法,如NPS基于预测值与探测值的相差比例的大小来排除恶意节点和PIC中基于三角不等性原理检测恶意节点等,均只能进行粗略的检测,有时甚至会影响预测的精度.网络的动态性和异构性等复杂特征使得恶意节点的攻击形式多样化,使得对网络坐标安全性的研究极其复杂.为此,Kaafar等人[60,64,65]在最近几年内进行了尝试,他们将恶意节点的攻击分为扰乱(disorder)、隔离(isolation)、排斥(repulsion)、系统控制(system control)4 种情形来考察坐标系统的安全性,并提出利用检查员基础设施(surveyor infrastructure)来检测恶意节点的行为,这种方法在一定程度上增强了坐标系统的安全性.

网络坐标的安全性研究,目前还处于研究的初级阶段,以前的研究更多地关注于方法的计算效率、精度和收敛性等方面的研究,而真正等到需要部署运用网络坐标系统的时候才发现安全性研究的重要性,由于坐标安全性研究的复杂性和必要性,决定了对此方面的研究将是一个长期的过程.

(5) 系统的实际部署研究

如何将已有的预测方法作为一项基础设施或者服务添加到具体的应用中去,即研究距离预测系统的实际部署问题.这是从理论研究走向实际应用的过程,它涉及到系统的一致性、可扩展性、稳定性、收敛性、安全性[65]、精确性以及存储、计算、通信开销等一系列问题.对此,网络坐标计算方法NPS[31,47,59]和非坐标距离预测方法Turbo King[23,43]都对此进行了研究,然而它们作为实际应用的初步尝试,并未达到能够满足应用需求的效果,可见实现距离预测系统以及研究如何部署将是目前以及今后研究的主要方面,也是网络距离预测技术从理论走向实践的必经之路.

6 结束语

网络距离预测技术研究作为新兴的热点研究领域,从早期非坐标预测方法到网络坐标计算方法、从集中式地标计算到分布式地标计算、从欧氏空间嵌入到复杂的非欧氏空间嵌入、从单纯的理论研究到实际系统的有效部署、从单一的度量标准到面向应用的度量标准的提出,可以说,网络距离预测技术的研究是一个非常活跃的方向.从整体上讲,目前在网络距离预测技术方面的研究还不够成熟,尚未建立起一套完整的理论体系和方法体系,而且从技术理论的完善到预测系统可靠的部署应用还距离甚远.

本文回顾了近年来学术界在网络距离预测技术研究领域的主要研究成果,在一个给定的研究框架下,指明了研究的主要方面以及各方面涉及的关键技术,并依据不同的标准加以分类,在分类的基础上分析、比较了各种具有代表性的预测方法,最后对各种精确性度量标准进行总结并指出了未来研究的趋势.

References:

[1] Kubiatowicz J, Bindel D, Chen Y, Czerwinski S, Eaton P, Geels D, Gummadi R, Rhea S, Weatherspoon H, Weimer W, Zhao B. OceanStore: An architecture for global-scale persistent storage. ACM SIGPLAN Notices, 2000,35(11):190−201.

[2] Maymounkov P, Mazieres D. Kademlia: A peer-to-peer information system based on the XOR metric. In: Proc. of the 1st Int'l Workshop on Peer-to-Peer Systems (IPTPS 2002). Berlin: Springer-Verlag, 2002. 53−65.

[3] Castro M, Druschel P, Kermarrec A M, Nandi A. SplitStream: High-Bandwidth content distribution in a cooperative environment. In: Proc. of the 19th ACM Symp. on Operating Systems Principles (SOSP 2003). New York: ACM Press, 2003.

[4] Hefeeda M, Habib A, Botev B, Xu D, Bhargava B. PROMISE: Peer-to-Peer media streaming using CollectCast. In: Proc. of the 11th ACM Int'l Conf. on Multimedia. 2003. 45−54. http://www.cs.sfu.ca/~mhefeeda/Papers/mm03.pdf

[5] Kostic D, Rodriguez A, Albrecht J, Vahdat A. Bullet: High bandwidth data dissemination using an overlay mesh. ACM SIGOPS Operating Systems Review, 2003,37(5):282−297.

[6] Zhao BY, Huang L, Stribling J, Rhea SC, Joseph AD, Kubiatowicz JD. Tapestry: A resilient global-scale overlay for service deployment. IEEE Journal on Selected Areas in Communications, 2004. 41−53.

[7] Ratnasamy S, Francis P, Handley M, Karp R. A scalable content-addressable network. In: Proc. of the ACM SIGCOMM. New York: ACM Press, 2001. 149−160.

[8] Stoica I, Robert M, Liben-Nowell D, Karger, Kaashoek MF, Dabek F, Balakrishnan H. Chord: A scalable peer-to-peer lookup protocol for Internet applications. 2001. http://pdos.csail.mit.edu/papers/chord:sigcomm01/chord_sigcomm.pdf

[9] Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In: Proc. of the 18th IFIP/ACM Int'1 Conf. on Distributed Systems Platforms (Middleware 2001). Berlin: Springer-Verlag, 2001. 329−350.

[10] Malkhi D, Naor M, Ratajczak D. Viceroy: A scalable and dynamic emulation of the butterfly. In: Proc. of the 21th Annual ACM Symp. on Principles of Distributed Computing (PODC 2002). New York: ACM Press, 2002. 183−192.

[11] Rewaskar S, Kaur J. Testing the scalability of overlay routing infrastructures. 2004. http://www.pamconf.org/2004/papers/262.pdf

[12] Chu Y, Rao SG, Seshan S, Zhang H. A case for end system multicast. IEEE Journal on Selected Areas in Communications. 2002, 20(8):1456−1471.

[13] Liebeherr J, Nahas M, Si W. Application-Layer multicasting with Delaunay triangulation overlays. IEEE Journal on Selected Areas in Communications. 2002,20(8):1472−1488.

[14] Awerbuch B, Shavitt Y. Topology aggregation for directed graphs. IEEE/ACM Trans. on Networking, 2001,9(1): 82−90.

[15] Jamin S, Jin C, Kurc A R, Raz D, Shavitt Y. Constrained mirror placement on the Internet. In: Proc. of the IEEE INFOCOM. Piscataway: IEEE Press, 2001

[16] Theilmann W, Rothermel K. Dynamic distance maps of the Internet. In: Proc. of the IEEE INFOCOM. Piscataway: IEEE Press, 2000.

[17] Ratnasamy S, Handley M, Karp R, Shenker S. Topologically-Aware overlay construction and server selection. In: Proc. of the IEEE INFOCOM. Piscataway: IEEE Press, 2002.

[18] Winter R, Zahn T, Schiller J. Topology-Aware overlay construction in dynamic networks. In: Proc. of the IEEE Int'l Competition Network ICN. 2004. http://cst.mi.fu-berlin.de/publications/pdf/rlm_cnpa_icn2004_final.pdf

[19] Wang W, Jin C, Jamin S. Network overlay construction under limited end-to-end reachability. In: Proc. of the IEEE INFOCOM. Piscataway: IEEE Press, 2005.

[20] Francis P, Jamin S, Jin C, Jin Y, Raz D, Shavitt Y, Zhang L. IDMaps: A global internet host distance estimation service. IEEE/ACM Trans. on Networking, 2001,9(5):525−540.

[21] Chen Y, Lim KH, Katz RH, Overton C. On the stability of network distance estimation. ACM SIGMETRICS Performance Evaluation Review, 2002. 21−30.

[22] Gummadi KP, Saroiu S, Gribble SD. King: Estimating latency between arbitrary Internet end hosts. In: Proc. of the 2nd ACM SIGCOMM Workshop on Internet measurement. New York: ACM Press, 2002. 5−18.

[23] Leonard D, Loguinov D. Turbo king: Framework for large-scale internet delay measurements. In: Proc. of the IEEE INFOCOM. Piscataway: IEEE Press, 2008.

[24] Srinivasan S, Zegura E. M-Coop: A scalable infrastructure for network measurement. In: Proc. of the 3rd IEEE Workshop on Internet Applications. Washington: IEEE Computer Society, 2003. 35−39.

[25] Wong B, Slivkins A, Sirer EG. Meridian: A lightweight network location service without virtual coordinates. In: Proc. of the ACM SIGCOMM. New York: ACM Press, 2005.

[26] Sharma P, Xu Z, Banerjee S, Lee SJ. Estimating network proximity and latency. ACM SIGCOMM Computer Communication Review. 2006. 39−50. http://networking.hpl.hp.com/s-cube/nv.pdf

[27] Guyton JD, Schwartz MF. Locating nearby copies of replicated Internet servers. In: Proc. of the ACM SIGCOMM. New York: ACM Press, 1995. 288−298.

[28] Ng TS, Zhang H. Predicting Internet network distance with coordinates-based approaches. In: Proc. of the IEEE INFOCOM. Piscataway: IEEE Press, 2002.

[29] Lim H, Hou JC, Choi CH. Constructing Internet coordinate system based on delay measurement. In: Proc. of the 2nd ACM SIGCOMM Workshop on Internet measurement. New York: ACM Press, 2003. 129−142.

[30] Tang L, Crovella M. Virtual landmarks for the Internet. In: Proc. of the ACM SIGCOMM Conf. on Internet measurement (IMC). New York: ACM Press, 2003. 143−152.

[31] Ng TS, Zhang H. A network positioning system for the Internet. In: Proc. of the USENIX Annual Technical Conf. Boston: USENIX Association Press, 2004.

[32] Waldvogel M, Rinaldi R. Efficient topology-aware overlay network. ACM SIGCOMM Computer Communication Review. 2003. 101−106. http://conferences.sigcomm.org/hotnets/2002/papers/waldvogel.pdf

[33] Pias M, Crowcroft J, Wilbur S, Harris T, Bhatti S. Lighthouses for scalable distributed location. In: Proc. of the Int'l Workshop on Peer-to-Peer Systems (IPTPS 2003). Berlin: Springer-Verlag, 2003

[34] Zhang R, Hu YC, Lin X, Fahmy S. A hierarchical approach to Internet distance prediction. In: Proc. of the 26th IEEE Int'l Conf. on Distributed Computing Systems. IEEE Computer Society, 2006.

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

[36] Shavitt Y, Tankel T. On the curvature of the Internet and its usage for overlay construction and distance estimation. In: Proc. of the IEEE INFOCOM. Piscataway: IEEE Press, 2004.

[37] Shavitt Y, Tankel T. Hyperbolic embedding of Internet graph for distance estimation and overlay construction. IEEE/ACM Trans. on Networking, 2008: 25−36. http://www.eng.tau.ac.il/~tankel/pub/TonHyp07.pdf

[38] Dabek F, Cox R, Kaashoek F, Morris R. Vivaldi: A decentralized network coordinate system. In: Proc. of the ACM SIGCOMM. New York: ACM Press, 2004. 15−26.

[39] Costa M, Castro M, Rowstron R, Key P. PIC: Practical Internet coordinates for distance estimation. In: Proc. of the 24th IEEE Int'l Conf. on Distributed Computing Systems (ICDCS). 2004. 178−187. http://ieeexplore.ieee.org/xpls/abs_all. jsp? arnumber=1281582

[40] Mao Y, Saul LK, Smith JM. IDES: An Internet distance estimation service for large networks. IEEE Journal on Selected Areas in Communications, 2006,24(12):2273−2284.

[41] Lehman L, Lerman S. PCoord: Network position estimation using peer-to-peer measurements. In: Proc. of the 3rd IEEE Int'l Symp. on Network Computing and Applications. 2004. 15−24.

[42] Wei L, Lerman S. A decentralized network coordinate system for robust internet distance. In: Proc. of the 3rd Int'l Conf. on Information Technology: New Generations (ITNG 2006). IEEE Computer Society, 2006.

[43] Xing C, Chen M. HNDP: A novel network distance prediction mechanism. In: Proc. of the IFIP NPC. 2007. http://www.springerlink.com/content/9k83022708m81336/

[44] Lua EK, Griffin T, Pias M, Zheng H, Crowcroft J. On the accuracy of embeddings for Internet coordinate systems. In: Proc. of the ACM SIGCOMM Conf. on Internet Measurement (IMC). New York: ACM Press, 2005.

[45] Hotz S. Routing information organization to support scalable interdomain routing with heterogeneous path requirements. 1994.

[46] Madhyastha HV, Anderson T, Krishnamurthy A, Spring N, Venkataramani A. A structural approach to latency prediction. In: Proc. of the ACM SIGCOMM Conf. on Internet Measurement (IMC). New York: ACM Press, 2006. 99−104.

[47] Madhyastha HV, Isdal T, Piatek M, Dixon C, Anderson T, Krishnamurthy A, Venkataramani A. iPlane: An information plane for distributed services. 2006.

[48] Freedman MJ, Lakshminarayanan K, Mazieres D. OASIS: Anycast for any service. In: Proc. of the 3rd Conf. on Symp. on Networked Systems Design & Implementation (NSDI 2006). Berkeley: USENI Press, 2006.

[49] Yan C, Randy K. Tomography-Based overlay network monitoring. sites. 2001. http://www.cs.berkeley.edu/yanchen/wnms

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

[51] Onbilger O K, Chen S, Chow R. A peer-to-peer network positioning architecture. In: Proc. of the 12th IEEE Int'l Conf. on Networks (ICON). Piscataway: IEEE Press, 2004.

[52] Zheng H, Lua EK, Pias M, Griffin TG. Internet routing policies and round-trip-times. In: Proc. of the 6th Int'l Workshop on Passive And Active Network Measurement (PAM 2005). Berlin: Springer-Verlag, 2005.

[53] Zhang B, Ng TS, Nandi A, Riedi R, Druschel P, Wang GH. Measurement based analysis, modeling, and synthesis of the internet delay space. In: Proc. of the 6th ACM SIGCOMM on Internet measurement (IMC). New York: ACM Press, 2006. 85−98.

[54] Wang G, Zhang B, Ng TS. Towards network triangle inequality violation aware distributed systems. In: Proc. of the 7th ACM SIGCOMM Conf. on Internet Measurement (IMC). New York: ACM Press, 2007. 175−188.

[55] Lee S, Zhang ZL, Sahu S, Saha D. On suitability of Euclidean embedding of Internet hosts. In: Proc. of the Joint Int'l Conf. on Measurement and Modeling of Computer Systems. 2006. 157−168. http://www.dtc.umn.edu/publications/reports/2006_32.pdf

[56] Eades P, Lin X. Spring algorithm and symmetry. Theoretical Computer Science, 2000. 379−405.

[57] Pietzuch P, Ledlie J, Seltzer M. Supporting network coordinates on PlanetLab. In: Proc. of the WORLDS. 2005. http://www.doc.ic.ac.uk/~peter/doc/nc-worlds05-camera1.pdf

[58] Ledlie J, Pietzuch P, Seltzer M. Stable and accurate network coordinates. In: Proc. of the 26th IEEE Int'l Conf. on Distributed Computing Systems (ICDCS). 2006. http://www.eecs.harvard.edu/~syrah/nc/icdcs06.pdf

[59] Ledlie J, Gardner P, Seltzer M. Network coordinates in the wild. In: Proc. of the 3rd Conf. on Symp. on Networked Systems Design & Implementation (NSDI 2007), 2007. http://www.eecs.harvard.edu/~syrah/nc/wild06-tr.pdf

[60] Kaafar MA, Mathy L, Barakat C, Salamatian K, Turletti T, Dabbous W. Securing Internet coordinate system: embedding phase. In: Proc. of the ACM SIGCOMM. New York: ACM Press, 2007.

[61] Ingwer B, Patrick G. Modern Multidimensional Scaling: Theory and Applications. Springer-Verlag. 1997.

[62] Stutzbach D, Rejaie R. Capturing accurate snapshots of the gnutella network. In: Global Internet Symp. 2005. 127−132. http://www.postel.org/gi2005/

[63] Stutzbach D, Rejaie R. Characterizing today's gutella topology. Technical Report, CIS-TR-04-02, University or Oregon, 2004.

[64] Kaafar MA, Mathy L, Turletti T, Dabbous W. Real attacks on virtual networks: Vivaldi out of tune. In: Proc. of the SIGCOMM Workshop on Large-Scale Attack Defense. 2006. 139−146. http://planete.inria.fr/dabbous/publis/lsad06.pdf

[65] Kaafar MA, Mathy L, Turletti T, Dabbous W. Virtual networks under attack: Disrupting internet coordinate systems. In: Proc. of the 2nd CoNext Conf. 2006. http://planete.inria.fr/dabbous/publis/conext06.pdf

王意洁(1971—),女,江苏镇江人,博士,教授,博士生导师,主要研究领域为网络计算,数据库,移动计算.

李小勇(1982－),男,硕士生,主要研究领域为网络计算,P2P 网络.