软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2011,22(1):84−100 [doi: 10.3724/SP.J.1001.2011.03935]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

互联网域间路由可扩展性*

张 威[1,3+], 毕 军[1,2,4], 吴建平[1,2,4]

[1](清华大学 计算机科学与技术系,北京 100084)
[2](清华大学 信息网络工程研究中心,北京 100084)
[3](武汉通信指挥学院,湖北 武汉 430010)
[4](清华信息科学与技术国家实验室(筹),北京 100084)

Scalability of Internet Inter-Domain Routing

ZHANG Wei[1,3+], BI Jun[1,2,4], WU Jian-Ping[1,2,4]

[1](Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)
[2](Network Research Center, Tsinghua University, Beijing 100084, China)
[3](Wuhan Communication Commanding Academy, Wuhan 430010, China)
[4](Tsinghua National Laboratory for Information Science and Technology (TNList), Beijing 100084, China)
+ Corresponding author: E-mail: zw@netarchlab.tsinghua.edu.cn

Zhang W, Bi J, Wu JP. Scalability of Internet inter-domain routing. *Journal of Software*, 2011,22(1):84−100.
http://www.jos.org.cn/1000-9825/3935.htm

Abstract: The problem of the scalability of inter-domain routing is of the main issues in the design of Next Generation Internet (NGI). This paper addresses the intrinsic essences of the problems created by the scalability of Internet routing by introducing the entropy routing information concept. Based on the theoretical model of entropy in routing information, potential solutions to the routing scalability problems are discussed, which focus on three different aspects with their respective benefits and limitations and their architectural evaluations on typical proposals. In the end, a conclusion on the challenges of the Internet routing scalability problem is drawn, and the direction of further research on this problem is explored.
Key words: inter-domain routing; next generation Internet; scalable routing; compact routing; architecture

摘 要: 互联网域间路由可扩展性问题是下一代互联网体系结构设计必须首先解决的关键问题之一.通过引入路由信息熵的概念,深入阐述 Internet 路由可扩展性问题的内在本质,并基于这一理论模型,分别从 3 个方面归纳解决路由可扩展性问题的 3 种可行思路.重点讨论了这 3 种思路应用于互联网路由系统的出发点和局限性.并就典型的具体提案从体系结构的角度进行了分析评价.最后总结路由可扩展性问题的挑战性,并展望了未来可扩展路由的研究发展方向.
关键词: 域间路由;下一代互联网;可扩展路由;紧致路由;体系结构

* 基金项目: 国家自然科学基金(61073172); 高等学校博士学科点专项科研基金(200800030034); 国家重点基础研究发展计划(973)(2009CB320501)
收稿时间: 2010-06-24; 定稿时间: 2010-08-27
CNKI 网络优先出版: 2010-11-05 11:50, http://www.cnki.net/kcms/detail/11-2560.TP.20101105.1150.003.html

© 中国科学院软件研究所 http://www.jos.org.cn

中图法分类号: TP393 **文献标识码**: A

Internet 日益增长的规模和应用需求对互联网技术的发展提出了巨大挑战.下一代互联网体系结构首先需要面对的问题是网络的大规模增长与提高路由性能的矛盾.IETF/IRTF 对互联网的路由可扩展问题给予了高度重视.该问题不仅引起了计算机理论、复杂系统等科研领域的兴趣,也是互联网产业界工程技术人员、设备制造商们最关心的问题之一.在下一代互联网体系结构设计的目标中,路由可扩展性不是唯一的要求,却是最重要的目标之一.

路由可扩展性研究一直都是路由体系结构设计的一个重要组成部分.2001 年,IRTF Routing Research Group (RRG)主席 Abha Ahuja 和 Sean Doran 决定成立工作小组 sub-group A,研究互联网新路由体系结构(inter-domain routing,简称 IDR)的需求问题,与此同时,瑞典的一个研究团队 Babylon 也独立开展了相似的工作.随后, Babylon 成为 RRG 下的另一研究小组 sub-group B,以区别于原有的 sub-group A,两个研究团队提出的对路由体系结构的需求有不同的认识,但在现有路由体系结构存在的可扩展性问题上达成了一致的基本共识[1,2].2006 年, IAB RAWS 的报告中全面描述了互联网路由可扩展性问题,并分析了产生该问题的根本原因.2007 年,根据报告的主要结论形成了 RFC4984[3],反映了业界对该问题的基本理解.

随后,该问题的研究大致可以分为两个方向:一个方向从分布式计算理论的角度探寻可扩展路由理论,并设计不同的路由算法;另一方向则主要从工程实现入手,试图设计新的互联网路由体系结构和协议解决路由可扩展性问题.两个方向都先后取得了阶段性的成果,但与彻底解决互联网路由可扩展性问题还有相当的距离.

从研究成果来看,理论研究方向重点研究可扩展路由算法,IRTF RRG 的 Future Domain Routing(FDR) Scalability Research Subgroup(RR-FS)进行了基础性的理论研究,从分布式计算理论出发,探寻了可扩展路由算法的理论,研究了大量针对路由可扩展性的紧致路由(compact routing)算法,并试图在实际 Internet 拓扑上寻求新的域间可扩展路由算法[4].但由于缺乏对 Internet 拓扑测量准确而全面的方法,同时难以构建精确的域间拓扑演化模型,暂时无法取得具有实际应用价值的 Internet 可扩展路由算法,该研究小组于 2007 年关闭.工程实践研究方向致力于从工程技术的角度全面分析和解决 Internet 路由体系结构的问题,具体来说,主要针对路由编址与路由机制如何解决支持包括移动(mobility)、多宿主(multihoming)和域间流量工程(traffic engineering)等功能在内的域间路由可扩展性问题.2007 年,Li 在文献[5]中提出了可扩展路由体系结构的十大设计目标.截止 2009 年 12 月,RRG 网站征集新提案 15 种,并列举原有 Active Proposals 9 种.ITRF RRG 试图在广泛征集新提案的基础上为 IETF 提供解决互联网路由可扩展性的建议,但目前各类提案分歧较大,还没有形成统一的解决方案或一致的解决思路.

针对 Internet 可扩展路由的研究,国内相关研究包括针对复杂网络的可扩展路由理论,张连明在文献[6]中综述了可扩展路由理论算法的模型与策略.同时,学术界密切关注当前国际研究的动向,崔勇在文献[7]中介绍了 IETF 对路由可扩展性问题的关注.其他针对解决可扩展性具体工程技术领域的文献包括:昝风彪等人在文献[8]中综述了主机标识协议,于士鹏在文献[9]中比较了与路由可扩展性密切相关的两种 ID/Locator 分离方案,徐小虎等人在文献[10]中介绍了基于虚拟聚合的解决路由可扩展性的方法.李继荣在文献[11]中则从路由选择、安全和拓扑 3 个方面介绍了不同技术对 Internet 可扩展性的影响.但目前研究还缺乏对互联网域间路由可扩展性问题的全面综述或基本理论分析.目前,国内在该领域的研究还相当有限,大都局限于路由可扩展性的几个相关领域,缺乏对该问题本质的全面揭示和理论分析.

本文试图从路由信息的本质剖析入手,揭示 Internet 路由可扩展性问题的基本内核,建立理论分析模型刻画影响域间路由可扩展性的 3 大关键因素,即地址空间规模、路由聚合度以及 AS 拓扑特性.继而归纳改进路由体系结构的 3 种解决思路,即从路由空间划分、提高路由聚合度和改进可扩展路由算法 3 个方面探索解决路由可扩展性问题的关键技术,并分析其应用于互联网域间路由的局限性.本文从体系结构研究的角度评述各类最近出现的典型提案,总结了众多不同技术方案的特点与局限,并最终据此归纳出解决互联网域间路由可扩展性问题存在的挑战性.该研究为设计未来互联网路由体系结构奠定了理论分析基础,并指明了努力的方向.

© 中国科学院软件研究所 http://www.jos.org.cn

本文首先介绍 Internet 路由可扩展性问题的背景和成因.第 2 节通过引入路由信息熵概念揭示路由可扩展性问题的本质,并给出互联网路由可扩展性问题的具体分析.第 3 节基于路由信息熵模型中的 3 个关键要素分析解决路由可扩展性问题的基本思路和关键技术,并对相关典型方案给予评价.第 4 节结合互联网当前的特点和演进的规律,讨论互联网可扩展路由体系结构面临的挑战性问题与难点.第 5 节展望未来的研究方向.

1 Internet 路由可扩展性概述

Internet 在向未来演进的过程中正面临着体系结构方面严重的挑战,其中一个突出的问题是网络规模不断扩张的现实需求与当前网络路由体系结构的矛盾日益尖锐.随着网络规模的大幅扩张和域间路由策略日益复杂(multihoming,域间流量工程等应用需求),全球 BGP 路由表急剧膨胀.有数据显示,2009 年,BGP 路由表项已超过 300 000 条,路由更新数量也相应地大幅增加.这种增长模式消耗了大量路由器存储资源,同时增加了维护路由更新的开销.该问题对提高互联网路由系统的效率和降低路由器开销带来了极大的影响.另一方面,IPv4 地址即将耗尽,如果采用更大的地址空间(如 IPv6)以满足未来网络规模增长的需求,并继续沿用现行的路由机制,势必导致路由表极度膨胀和路由更新大幅增加,最终可能成为下一代互联网顺利演进的主要障碍之一.

根据 2006 年 IAB 的分析报告[3],路由表膨胀导致全球 BGP 路由更新持续增加,而大量路由更新(BGP update churn)消耗了大量的网络带宽和路由计算资源.网络监测数据显示,BGP 路由表尺寸和更新报文规模保持加速增长的总体趋势.可以设想,当更大的地址空间(IPv6)进入路由系统后,这个问题会十分严重.更大的路由表和转发表直接增加了路由器设备的成本和网络运营商的开销.2008 年,粗略的研究初步估算,BGP 每宣告一条路由表项,增加的经济成本是 0.04 美元/年.按照当时的路由器数量和路由表规模,全球 BGP 路由表开销约 20 亿美元左右.随着网络规模的增加,这一成本将持续递增.IAB 报告中指出,Internet 大规模路由体系结构最紧迫的两个问题是路由可扩展性和 IP 功能语义重载问题.

路由表膨胀的成因是多方面的,其中比较一致认同的结论有 3 个方面:一是多数 ISP 选择 Provider Independent(PI)地址,这样做可以保证在更换服务提供商时避免了更换网络地址(renumbering),从而保证系统管理员不必重新配置防火墙等 IP 地址相关的设备.但这一策略导致服务提供商(provider)必须单独宣告客户(customer)的 PI 地址,从而在网络中形成不可聚合的路由前缀.PI 地址的使用在一定程度上影响了 IP 地址的聚合性;二是网络中的多宿主 multihoming 连接,多宿主提高了连接的可靠性,还带来了优化接入成本、增加网络可靠性和流量均衡等好处.但这种连接方式仍然需要 provider 在 BGP 中宣告不可聚合的 customer 地址,同样破坏了路由地址的可聚合性;三是域内的流量工程(traffic engineering),由于路由采用最长前缀匹配的基本策略,为了实现域内的流量工程,影响域间引入流量的路径,往往需要向外广播掩码更长的子前缀,这种方式明显地会导致路由前缀碎片的进一步增多,同样不利于路由聚合.按照目前的路由机制,域间前缀策略表达必须以牺牲路由可聚合性为代价,而且这一影响会扩散到全局 BGP 的范围,最终导致了 Internet 路由严重的可扩展性问题.

IAB 归纳的上述结论揭示了 Internet 路由系统可扩展性问题的直接成因,但究其问题的根本则在于现行路由体系结构的缺陷.概括来说,IP 路由机制形成于 20 世纪 70 年代,其网络规模和路由特性都无法和当前的 Internet 应用相提并论.在可扩展性方面,历史上有两次意义重大的技术改进很大程度上缓解了网络规模增长的需求:一是采用 CIDR 编址结构,提高了 IP 地址的利用率;另一改进是采用网络地址翻译(NAT)技术对部分私有 IP 地址复用.但这两种技术都仅在一定程度上缓解了网络规模与 IP 地址空间的矛盾,并未对路由机制进行实质改进.随着 IPv6 地址的分配和初步部署的启用,地址空间的问题可以得到根本性解决,然而建立在更大规模地址空间上的可扩展路由机制显得更为紧迫.

设计下一代互联网高效的大规模路由机制,必须深刻理解路由可扩展性问题的本质.为便于分析,我们对当前域间路由模式进行适当而合理的抽象,建立基于路由信息熵的理论模型.

2 基于路由信息熵的域间路由可扩展性理论分析

在分布式系统的路由体系中,节点通过相互交换路由信息实现拓扑发现功能.通常,我们将网络路由过程抽

© 中国科学院软件研究所 http://www.jos.org.cn

象为建立在加权图 $G(V,E)$上的分布式路由算法,其中,V 是节点,E 是连接节点的边.在这个过程中,节点间的拓扑关系是基本制约条件.路由算法依据某种策略,通过路由信息(包括报文携带的相关信息和路由表信息),最终确保每个节点能做出一致的转发决策(转发路径收敛,避免出现路由环路或路由失败).在分布式路由过程中,路由信息的交互起到了决定性作用.我们将围绕路由信息这一基本要素深入分析域间路由模式在可扩展性方面的核心地位.

在建立理论分析模型之前,我们首先借用信息论的基本概念描述路由过程所需的信息量.Shannon 在文献[12]中将信号信息量的大小等价于“不确定性”的度量,并建立了香农熵的定义.该定义将信息中关于信息度量和缺乏信息条件下的不确定性联系起来,揭示了信息能消除选择(判决)不确定性的本质特性.从信息论的观点来看路由过程,节点转发分组同样存在路由不确定性选择问题,而路由信息是消除转发不确定性的依据.借鉴信息论中的信息熵来量化路由信息,我们定义“路由信息熵”是节点对路由转发不确定性的度量.如果用路由信息熵量化,每个节点转发分组时(包括到该节点自身的),不确定性之和构成了该节点路由信息的总量.只有当网络路由协议满足所有节点路由信息量的存储和交换条件时,才能完全消除转发的不确定性,并保证路由过程的正确实施.否则,在路由信息不足以覆盖全部路由信息熵时,存在转发的不确定性,分布式算法不能保证选取最优路径或者路由成功.

路由信息熵的计算采用信息论的基本方法,即对不确定性进行统计意义的度量.假设网络节点 $i(i\in V)$的度是 d_i,如果网络节点编址信息不包含拓扑信息,不失一般性,可以假设转发是等概率的,用 P_i 表示.此条件下有最大不确定性,熵值最大,那么有 $P_i=1/d_i$.

转发到任一目的节点的不确定性的数学期望,即到任一目的节点的信息熵 H 为

$$H=-\sum_{d_i}P_i\log_2 p_i=\log_2 d_i.$$

对于有 n 个目的节点的网络,节点 i 需要 n 条路由表项确保正确转发到所有目的地址的分组.路由信息总量的信息熵表示为 E_n:

$$E_n=nH=n\log_2 d_i \tag{1}$$

由公式(1)可知,在分布式系统中,为了完全消除路由的不确定性,网络节点所需路由信息量随目的节点数 n 和该节点的度 d_i 的增加而增加,我们将这个度量定义为路由信息熵.路由信息熵的大小是对路由不确定性的描述,我们用它来揭示网络的路由可扩展性.如果各网络节点不能存储覆盖全部路由信息熵的信息量,或者网络不能及时更新这些路由信息,则会导致路由性能下降.路由可扩展性问题是指路由性能随网络规模的增长而下降的问题,具体对互联网域间路由而言,指的是当网络设备和路由算法无法保证路由信息正常存储交换时,路由性能受到何种程度影响的问题.

$$E_n=n\log_2 d_i=\sum_j A_j\log_2 d_i\approx\frac{A}{\rho}\log_2 d_i \tag{2}$$

对于互联网域间路由的场景,需要作具体的分析.根据 BGP 路由协议,将 AS 抽象为转发节点的基本单位,而路由转发信息的粒度则是基于 IP 前缀的.在公式(1)中,d_i 表示 AS 的度,而 BGP 中 IP 前缀的数量决定了 n 的大小.IP 前缀表示路由地址空间中连续的一块(A_j),如果用 A 表示直接可路由地址空间的大小(IPv4 中上限为 2^{32} 而 IPv6 中上限是 2^{128}),ρ表示地址空间的平均聚合因子(IP 前缀的平均聚合程度),则它们的比值会直接影响网络的路由规模.公式(2)中的路由信息熵表示了 BGP 路由可扩展性方面的基本特征.

根据上述理论模型,在当前 BGP 路由系统中,影响路由信息熵的要素有 3 个:一是网络节点的规模(可路由地址空间 A),二是路由可聚合性(平均聚合因子ρ),三是网络的基本拓扑特性(节点度 d_i 的分布).首先,地址空间直接决定了路由规模的大小.随着网络的发展,地址空间的递增是刚性需求,“可路由的”互联网地址空间总是趋向于持续的递增.从互联网发展的历史来看,网络规模持续递增,且其规模增长不会受到有限地址空间的约束;其次,聚合因子反映了路由信息可聚合性的基本特征.当路由系统中的 IP 前缀表现出较差的聚合性时,聚合因子ρ较小,路由信息熵较大,可扩展性差.互联网是大规模的自组织系统,缺乏统一的网络规划.网络被划分为多个独立的管理边界,其有序性无法保障.PI 地址的大量使用导致了地址聚合上的困难;第三,连接度 d_i 高的 AS 路由信

© 中国科学院软件研究所 http://www.jos.org.cn

息熵较大.互联网拓扑测量显示[13],自治域(AS)级网络拓扑体现出服从幂律分布的“小世界”特性.在这种拓扑结构中,AS 连接度的分布极不均衡,呈现出无标度网络的基本特点,少数 AS 连接度极高.这种不均衡的网络拓扑演化规律,是互联网演进的基本特征之一.连接度高的节点更容易面临路由可扩展性的问题,换句话说,不同拓扑位置的节点对解决路由可扩展性问题的迫切程度和重要性均不相同.路由信息熵模型客观地反映了影响路由不确定性的关键因素及其本质联系,它是我们分析路由可扩展性问题的基本理论模型.

根据上述分析,路由信息熵表现为一个持续熵增的过程.路由信息膨胀越来越严重,最终成为影响路由可扩展性的症结.事实表明,互联网路由信息的总量随时间的推移而递增.由于 BGP 路由协议的路由信息完全由路由表结构承载,因此,可扩展性问题最直接地表现为 BGP 路由表膨胀和路由更新开销增大.综上所述,总体路由信息熵的增长导致分布式系统中路由转发设备需要存储的路由信息量急剧增加,从而构成了互联网域间路由可扩展性问题的基本内核.

3 解决域间路由可扩展性的基本思路、关键技术及其典型方案评价

解决域间路由可扩展性问题就是使每个节点存储的路由信息量不随网路路由规模的增长而相应地快速增加;同时,尽可能地使路由性能不受到较大影响.针对路由信息熵理论模型分析的 3 个关键因素,从工程上解决互联网域间路由可扩展性问题大致可归纳为 3 种基本思路:一是“分治”的思想,划分不同的路由空间限制地址空间的规模 A;二是改进路由聚合技术,压缩路由信息,增大平均聚合因子 ρ;三是改进路由算法,例如,在编址中嵌入路由信息,或在报文中携带部分路由信息,而不是仅依赖节点分布式存储和交换路由表信息.编址和报文中的路由信息使得节点转发报文到每个端口时的不确定性不再是等概率的(等概率不确定时熵有最大值),因此可以达到减小路由信息熵的目的.

上述 3 种思路都有若干典型的路由方案和相应的关键技术,但这些方案还不足以彻底解决路由可扩展性问题,因而表现出不同的局限性.

3.1 “分治”的思想与划分路由空间技术

第 1 种思路是采用分治的思想,即划分不同的路由空间,不同的路由空间采用各自的路由编址,相互不交换路由信息,跨越路由空间时需要进行地址映射.由于每个路由空间的规模是有限的,因此路由信息熵的大小得到控制,缓解了网络的可扩展性问题.

最典型的划分路由空间思想在核心边缘分离(core edge separation,简称 CES)方案中得以体现.这一类方案针对互联网现有的运营规模、拓扑特性和商业策略决定了两类不同特性的网络.即为数不多的核心传输骨干网络(transit core)和大量处于边缘不提供穿越服务的网络(edge networks).支持分离方案的研究者认为,核心网络(transit core)表现出的特点有:在 AS 级拓扑上连接度极高,商业模式上无 Provider.这类网络总数量有限,规模较小,拓扑相对稳定.Massey 等人在文献[14]中测量的结果表明,核心 AS 增长的速度只有边缘网络(edge networks)的约 20%.边缘网络则相反,AS 级拓扑连接度较小,与拓扑邻居商业关系复杂(provider,peer 和 customer 等).整个 Internet 的拓扑结构表现为一个密集连接的穿越核心周围伴有大量的边缘网络.在这样的拓扑结构上建立不同的路由空间,符合可扩展路由的要求.最自然的划分就是将核心网络和边缘网络划分到不同的路由空间,边缘网络间的路由可能需要穿越核心网,而边缘网络内的路由采用不同的路由机制.

如果将网络划分为核心路由区域和边缘路由区域,则边缘路由区域通过核心路由区域提供穿越路径,穿越核心的路由对边缘是透明的,核心也不关心边缘的路由状态.这一结构较符合互联网的域间拓扑特性.因为互联网有一个连通度密集的穿越核(transit core),结构较为稳定,而为数众多的边缘网络则连通度较小,拓扑结构动态性较强.这两种不同拓扑特性的网络可以分别采用不同的路由模式,例如,在拓扑关系比较稳定的核心网络采用 PA (provider aggregateable)地址提高路由的可聚合性,而边缘网采用 PI(provider independent)地址支持网络的迁移(portability)能力,即当网络改变接入位置时无需重编址(renumbering).

具体方案又可以分为“仅在控制平面划分路由空间”和“同时在控制和转发平面划分路由空间”两类.前者比较典型的方案有“链路状态-路径向量混合路由协议(hybrid link state path vector routing protocol,简称

© 中国科学院软件研究所 http://www.jos.org.cn

HLP)”[15],而后者包括 LISP[16],Ivip[17],Apt[18],TIDR[19]等众多 core-edge 分离(CES)方案.其中:HLP 仅在控制面改进和拓展 BGP 协议功能;而 CES 类方案则在控制面和数据转发面彻底采取路由地址空间分离的策略,需要部署额外的网络设施,如映射系统服务器、隧道封装入口和出口路由器(ITR/ETR)等网络设备.

3.1.1 HLP

HLP 引入了链路状态路由模式建立“AS 的联合体路由区域(AS group)”,如图 1 所示.图中双箭头表示 AS 间的 peer 关系,单箭头则表示 provider/customer 关系.首先,在全网划分 AS group,AS group 内部用类似链路状态(LS)模式交换路由信息,AS group 间则延用原路径向量路由机制.该方案实际上对互联网域间路由划分了新的路由信息交换区域,在 AS group 之间可以屏蔽部分 AS 间的拓扑更新动态.在联合体路由区内部的链路状态路由,则建立在 AS 级的粒度而不是路由器级的粒度上.这种改进一方面限制了域间拓扑更新对全网的影响,另一方面并不影响路由策略的表达和数据转发过程.HLP 仅从控制平面人为地划分路由信息交换的区域,也可以理解为在控制平面通过划分不同的路由空间实现两个路由层次.该方案不影响数据转发平面,只需要扩展域间路由协议的功能.

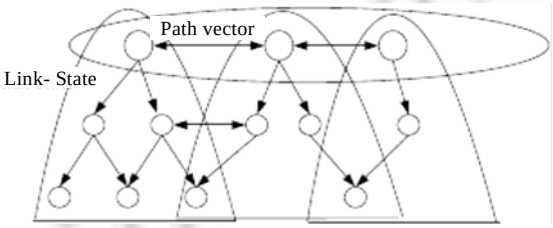

Fig.1 HLP routing

图 1 HLP 路由模型

但该方案的局限也很明显:首先,是否总能找到合适的 AS group 路由区划分是一个部署上的难题.AS 接入互联网是独立决策的行为,没有任何统一规划域间层次化关系的机构和机制,这种宽松的环境是互联网本质的特点.如果能够统一协调层次结构,则很容易做到路由聚合;其次,该方法对参与路径向量协议的 AS 并没有带来可扩展性方面的收益.因为 AS group 内部的地址空间仍然有可能是不可聚合的,路由前缀的规模也有可能并没有下降;第三,决定路由层次的 AS 间关系(provider-customer,peer to peer)有可能发生改变,如某个 AS 需要加入路径向量路由空间成为超级路由区成员,则可能会引起超级路由区整体结构的变动,需要重新构建新的超级路由区.而对于这种路由空间划分的调整,域间协议无法自动配置,需要多方协商,涉及到多个 AS 的重新配置.

3.1.2 核心边缘分离类(CES)方案

相比于 HLP,CES 类方案是更加彻底的“路由空间分离”:在数据转发平面也划分不同的路由地址空间,不同的地址空间不能互通,穿越路由空间时必须进行地址空间映射,路由系统中需要增加地址映射服务功能.CES 类方案把互联网人为地划分为传送核心网络和边缘网络,在 IETF RRG 2010 年网站统计的方案中,属于核心边缘分离(core edge separation,简称 CES)类型的方案有 LISP[16],Ivip[17],Apt[18],TIDR[19], IPvLX[20]等.其中,较完备和详尽的典型方案是 LISP 和 Ivip.

CISCO 提出的 LISP 的分离方案将互联网划分为核心路由区和边缘网络两类路由空间.通过 ITR(ingress tunnel router)和 ETR(egress tunnel router)实现边缘网络穿越核心网络的功能.边缘网进入核心网络前由 ITR 查询映射系统,获得目的地址对应的 ETR,然后对分组封装,通过 IP 隧道转发到 ETR.ETR 解封装后,将报文转发到目的地址.这个过程需要 ITR/ETR 保持正确的地址映射关系,映射系统是独立于路由转发功能的新增功能,成为设计的重点.LISP 在映射系统的实现上曾分别提出了 ALT,DHT,CONS 和 NERD 这 4 个版本.为了实现增量部署和平滑过渡,LISP 还设计了 ITR/ETR 代理等新功能.LISP 专门开发了映射请求和应答协议,允许 ITR 在发送第 1 个包时进行 ETR 探测,以减小映射带来的时延.

Ivip[17]的设计与 LISP 具有类似的框架,主要区别在于映射的实现方式和穿越核心网络的技术细节不同.例

© 中国科学院软件研究所 http://www.jos.org.cn

如,Ivip 为了避免隧道带来的路径 MTU 探测(PMTUD)问题,采取了地址重写技术.分组进入 ITR 后,ITR 将分组头的源目的地址改为 ITR/ETR 地址,同时利用 IPv4 和 IPv6 分组头中未使用的位携带边缘网络源、目的地址.该技术在 IPv4 中称为 ETR Address Forwarding(EAF),而在 IPv6 中称为 Prefix Label Forwarding(PLF).该方法避免了分组由于隧道封装而变长最终超出 MTU 限制的问题,使得整个端到端的转发路径上 MTU 是一致的.

在 CES 类方案中,根据路由信息熵模型,核心网络的可路由地址空间决定了路由信息熵的大小.如果设定全局路由空间的规模为编址空间的 50%,则路由信息熵减小为一半左右,模拟研究的结果表明,这时,CES 方案能够有效地减少约 50%的路由表项.与此同时,核心网络的增长速度和动态性远低于边缘网络.因此,核心网络的全局可路由空间能够保证网络规模增长的需求.另外,在核心网络或边缘网络,还可以根据条件进行可聚合的层次化编址而不影响其他路由空间.

CES 类方案主要存在如下局限性:首先是如何划分核心网络和边缘网络的边界问题.现有的方案缺乏明确的界定,造成实际部署的困难;其次,在数据转发平面分离地址空间的方式破坏了通信端到端的原则.例如,由于源目的地址在转发过程中发生改变(隧道封装后源目的地址都在 payload 中),网络防火墙的配置和性能以及安全验证等机制在核心网受到很大影响;再次,隧道封装增长了分组的长度,导致出现 PMTUD 问题,并增加了协议开销;最后,封装和地址翻译带来额外的网络开销,映射查询可能会增大网络时延.

另外,CES 类方案必须解决映射系统本身的问题.当前,映射系统的设计主要考虑和权衡 3 个方面的问题:一是考虑到映射查询的可扩展性问题,采用层次化的映射系统结构比较理想,但查询时效性较差,一个查询需要经过若干次递归查询,可能会产生较大延时,也可能对新发起的流不利.而映射系统采用集中式或扁平的分布式结构则可扩展性差,容易产生瓶颈效应;二是考虑到提高映射效率可以采用 Cache 机制,DNS 系统有效地采用多层的缓存机制构建了域名解析服务.但是,ID/Locator 映射查询不仅节点规模比 DNS 的域名大很多,而且其动态性也比 Full Qualified Domain Name(FQDN)的变更频率要高.如果 Cache 的层次太多,则更新成本高且容易产生错误解析.减少 Cache 的层次和容量又会降低查询效率;三是考虑映射系统的信息更新采用 Push 模式还是 Pull 模式或是 Push+ Pull 的混合模式.Push 模式对网络拓扑的变更反映灵敏,路由失效概率低,但网络开销较大,而且对映射系统的计算资源要求较高;相反地,Pull 模式是按需查询的方法,资源开销较小,但时效性差.而且如果映射 Cache 不能正确反映拓扑的更新状态,则容易产生路由失效;混合模式是对两种模式的权衡,试图在两者之间找到一个最佳平衡点.映射机制的研究还处于最初始的阶段,许多提案仅作为结合具体 ID/Locator 分离方案的配套部署.但最近,RRG 工作组已有部分研究人员开始重视映射的效率和可扩展性等问题.在 2009 年底提出的 15 个针对解决路由体系结构可扩展性的备选提案中,就有 3 个是专门针对映射机制的.映射机制需要考虑的问题远比想象的要复杂,其性能不仅和部署的网络位置有关,而且网络流量分布的特性也会对映射效能产生很大影响.另外,映射机制的安全性也十分重要.需要实际部署大规模映射系统加以验证各种机制的效能,帮助我们理解映射本身存在的问题.这方面的研究有待将来进一步加以深入.

采用划分路由空间的思想提供了最接近现有路由体系结构的方案,最大程度地利用现有协议实现和网络既有部署设施,依靠 BGP 协议扩展和特殊路由配置功能,在一定程度上缓解了 BGP 路由可扩展性问题,可以应对在新路由体系结构全面部署之前的现实问题,适宜应对最急迫的近期路由可扩展性需求.此类方案主要依赖于对路由空间规模的限制,达到减少 BGP 前缀数量的目的.

划分路由空间方法的局限性主要表现在:首先,不同路由空间之间无法直接路由,需要地址空间映射技术和全局地址映射系统的支持,边缘网络穿越核心网络时需要将一个地址空间的地址映射为另一地址空间的接入位置.由于映射本身的动态性,需要有一个映射系统提供服务支持.通常,映射的规模要求一个全网覆盖的分布式系统为所有路由空间边界的路由器提供地址映射;其次,划分地址空间的方法破坏了通信的端到端原则,而且需要隧道封装或地址翻译技术穿越不同的路由空间,由此而产生了额外的地址映射和地址转换开销;最后,核心/边缘网络分离还需要解决划分边界如何确定的问题.

3.2 路由信息压缩与路由聚合技术

第 2 种思想是,通过路由聚合对路由信息进行压缩.路由信息的压缩包括无损压缩和有损压缩两类方法:无

© 中国科学院软件研究所 http://www.jos.org.cn

损压缩针对有效压缩冗余路由信息,不影响路由性能;而有损压缩则以牺牲少量路由信息为代价(次优路由产生一定路径延展)来降低路由表尺寸.这一思路的基本出发点是,在路由表中提高转发信息的“粒度”,即提高路由聚合因子,以降低路由信息熵.

无损路由信息压缩考虑到路由表项在信息存储时的冗余,相关研究[21]表明:BGP 路由表中约有一半前缀是被覆盖的前缀(covered prefixes);而另一半是覆盖前缀(covering prefixes),即有大量子前缀存在.而这些大量有覆盖关系的前缀可能具有相同的转发属性(如,下一跳地址、转发端口或隧道封装)或路由属性(如,AS path 或 BGP attribute).通过适当归并具有相同转发属性的前缀可以压缩 FIB 表[22],而聚合相同路由属性的被覆盖前缀可以压缩 RIB 表.在路由信息压缩的技术中,通常将前缀定位到 patricia trie 树类似的数据结构,通过树结构的父子继承关系分析前缀的相互覆盖关系.相邻的树节点也可以合并成新的节点(新的聚合前缀).无损路由信息压缩不影响路由的转发路径,但会带来较大的计算开销.而且其压缩效能有限,在特殊情况下,当路由表中路由冗余信息有限时,则基本不能起到减小路由表尺寸的作用.在 BGP 协议中,信息冗余毕竟不是造成可扩展性问题的本质原因.总之,无损信息压缩的极限是路由信息的冗余部分,如果超过这个压缩范围,必将影响路由的性能,则成为有损压缩的方法.

有损压缩路由信息考虑通过适当牺牲有限的路由性能,以换取路由信息熵的大幅减少.有损压缩路由信息的技术,较典型的包括层次化路由方法、虚拟聚合(VA)[23,24]方法等.层次化聚合在保证路由可达的前提下,不一定每个转发节点都必须存储完整的到所有其他目的节点的全部最优路由信息,可以合理构建逻辑路由层次,允许每个节点仅对网络中的部分节点(如某一路由层次内的节点)有确知的最优路由信息,而对网络中其他节点(非本路由层次的远端节点)按路由层次聚合路由(类似电信网络的层次结构),并通过本路由层次选定的交换节点转发.分布式路由系统中具有代表性的层次路由模型是 *k-k* 模型[25].该模型假定任何路由层次上给定路由区域内的节点间路径仅在该区域内.类似的层次化路由方案还包括 Nimrod[26],ISLAY[27]等,每个节点的路由表大小主要与其所在层次的路由规模有关,只要限制每一层次的路由规模,则可以保证每个节点的路由表大小满足存储的要求.但层次路由引入了另一个问题,即路径延展(stretch),即部分路由不是最优路由.实际路径长度与最短路径长度的比是衡量路径延展的度量.在 *k-k* 的层次化路由模型中,路径延展度与网络的拓扑特性有密切联系.一般而言,层次化路由构建一个树形结构,划分 N 个节点到层次树的 m 层中,每层有若干个簇,每簇有 n 个节点.每簇构成的子网络的网络半径 d 与 n 有如下幂律函数关系:

$$d \leqslant bn^{v}+c \tag{3}$$

公式(3)中,b,v,c 是正参数,其中,$0 \leqslant v \leqslant 1$.而网络中的平均路径长度(跳数)$h$ 满足如下幂律函数关系:

$$h=aN^{v} \tag{4}$$

公式(4)中,a 是正参数.构建路由层次必须充分考虑这些参数的优化过程,最终,这些参数将对路径延展起到关键性的作用.如果定义 h_c 是层次路由的路径长度,h 为最优路径长度,则有路径延展 E 表示为

$$0 \leqslant \frac{h_c}{h}-1 \leqslant E \propto \frac{1}{a(N-1)N^{v}}\left\{n\left[b\frac{N^{v}-N^{v/m}}{N^{v/m}-1}+c(m-1)\right]-b\frac{N^{v+1}-N^{(v+1)/m}}{N^{(v+1)/m}-1}-c\frac{N-N^{1/m}}{N^{1/m}-1}\right\} \tag{5}$$

而此时,路由表压缩的比例 R 为

$$R=\frac{m\sqrt[m]{N}}{N} \tag{6}$$

在 *k-k* 模型中,拓扑连接稀疏的网络,公式(4)中的参数 v 介于 0~1 之间.在层次化路由模式下,平均路径延展较小.而对于拓扑连接密集的网络,平均跳数 h 较小,由公式(4)可知,v 趋于 0,因而在公式(5)中,延展度 E 很大.但是,互联网 AS 级拓扑具有无标度(scale free)网络的特性,有一个连通度密集的核心,节点间平均距离较小(实际测量平均 AS 距离 h 约为 3~4 之间).这时,在域间采用层次路由会导致产生较大的平均路径延展.尽管该模型没有明确规定构建层次的算法,并且关于平均路径长度 h 与 N 的幂律关系在互联网拓扑中也难以准确度量,Krioukov 等人仍然根据测量的数据和基本假定条件推导出互联网拓扑应用 *k-k* 模型的平均延展度在 10 倍以上[28],同时还无法保证最差延展的上限(unbounded worst stretch).另一方面,层次化聚合还要求网络编址采用层

© 中国科学院软件研究所 http://www.jos.org.cn

次可聚合路由地址(PA 地址),即要求网络编址符合路由层次的组织结构.但互联网路由体系结构已决定了路由管理边界的划分,而且每个独立的自治系统在缺乏全网统筹规划的情况下无法有效保证路由编址按路由层次组织.域间路由的可扩展性问题要求在域间构建逻辑路由层次的同时,必须实现路由编址的统一分配和管理机制.以上两方面决定了 *k-k* 层次路由模型不能满足域间可扩展路由的实际需要.

虚拟聚合(VA)技术采用了虚拟化的思想,回避了路由层次聚合在地址分配和部署上的局限,不一定要求可聚合的前缀在拓扑或路由层次结构上位于相同或相邻的物理拓扑位置(设备).因为层次的构建是“虚拟的”.虚拟聚合(VA)方法首先构建一个逻辑上的虚拟前缀(virtual prefix,简称 VP),它覆盖若干在地址空间上可聚合的 BGP 前缀.然后,VA 指定负责虚拟前缀的路由设备(aggregate point router,简称 APR),由 APR 在网络中宣告 VP 并维护被给定 VP 所覆盖的各子前缀的真实转发路径.由于 APR 仅宣告 VP,因而有效地屏蔽了被 VP 所覆盖的大量的 BGP 前缀碎片,网络中的路由聚合性得以加强.VA 可以同时在域内和域间部署 APR,在域内应用可以大幅减小路由器的 FIB 尺寸,但对 RIB 的大小不能产生影响.在域间逐步部署 VA 可以有效地增强 BGP 路由前缀的聚合度,当部署的规模达到一定程度时,可以预期,全球路由表高度聚合于少量的 VP.虚拟聚合实际上用虚拟的方法构建了路由层次(APR 可以组成一层或多层),但避免了 PA 地址的分配和统一管理问题.网络迁移(更换 ISP)仍可实现路由聚合而不需要网络重新编址(renumbering),但 VA 也与其他层次化路由方案一样,不可避免地引入了路径延展问题,同时还引入了网络流量的重定向问题.当 APR 部署的位置不合理时,有可能将大量流量引入某些 APR,从而导致路径拥塞.另外,APR 采用隧道封装技术转发分组,也破坏了端到端原则,并增加了路由的开销.

路由信息聚合是提高路由可扩展性的关键技术之一,可以应对中(长)期域间路由可扩展性的需求.此类方案强调路由聚类,通过改进路由可聚合程度来提高路由可扩展性.具体提案包括虚拟路由聚合和层次化编址聚合路由两类,前者典型的方案如 CRIO[24],后者包括一类通过 Provider Aggregateable(PA)地址聚合路由的层次化编址方案.

3.2.1 虚拟路由聚合方案

CRIO 是基于虚拟聚合(VA)的域间路由方案,采用了虚拟前缀(VP)聚合的方法来划分路由层次,该方法在 AS 间部署一个覆盖网络(overlay),由若干负责虚拟前缀的路由器(APR)和提供商边缘(PE)路由器组成,PE 路由器将边缘网络的前缀通过映射(mapping)或多跳 BGP 的方式告知相应的 APR.这时,APR 根据每条映射的路由前缀建立到各个 PE 路由器的隧道转发表项.同时,APR 通过 BGP 对全网(overlay)广播它所负责的 VP 及其自身的地址,每个站点都能学习到 APR 的地址和 VP 的转发信息,能够正确地将目的前缀映射到所属的 APR,APR 通过隧道转发分组到相应的连接边缘目的网络的隧道出口路由器(egress tunnel router,简称 ETR).这样,大量前缀碎片被 PE 路由器屏蔽,不再通过 BGP 在全网通告而仅由相应的 APR 负责.每个 APR 仅负责一个虚拟前缀覆盖下的前缀碎片,并维护到各前缀碎片的 PE 路由器隧道转发信息.VA 的思想还可以很方便地应用于域内压缩路由器转发表(FIB)的尺寸,但仅在一个 AS 内的 VA 部署不改变域间路由表(RIB)的规模.

CRIO 的方法仅依赖于现有的技术和部署配置条件就可以实现,并且可以满足增量部署的要求,基于虚拟聚合的技术无需对地址分配和编址有特殊的要求.VP 覆盖的路由前缀碎片越多,则聚合效率越高,而且不部署 CRIO 的地址前缀仍然可以照常参与 BGP 路由信息交互而不受影响.

CRIO 带来的主要问题是层次化路由的基本症结:路径延展.采用 APR 的隧道转发通常都不是最优路由,层次路由模型的理论已经证明,聚合程度越高,延展度越大.因此,VP 的选择也是在路由聚合程度与路径延展度间权衡的结果.相对次要的问题还有 VA 导致了流量的重定向,使大量流量集中到 APR.映射和隧道机制增加了路由的开销.APR 的位置对网络性能影响较大,需要全局优化.而这个优化问题是 NP 难问题,甚至很难使之具备全局优化的条件.最后一个问题是,CRIO 不能保证满足复杂而灵活的域间路由策略,而策略路由是运营商的基本经济模式决定的运营需求.

3.2.2 层次化编址路由聚合方案

采用层次化编址聚合路由的方案通常也被称为“移除(elimination)”方案,主张完全采用 PA 地址,将不可聚

© 中国科学院软件研究所 http://www.jos.org.cn

合的 PI 地址从网络中移除.“移除”方案为了保证路由的可聚合性,需要对网络中的节点进行拓扑相关编址或地址分配.网络地址与接入位置相关,当站点有多个接入位置(multihoming)时,不惜同时配置多个网络地址.在路由信息中,所有地址都是拓扑可聚合的,因此,路由聚合性得以大幅改善.属于“移除”类的典型方案包括 Shim6[29],SIX/ONE[30],NIRA[31],ILNP[32],GLI[33],HIP[34],hIPv4[35]以及 HAIR[36]等.

移除方案采用了拓扑相关的网络编址,当网络的接入位置发生变化时,地址也将改变.因此,在体系结构上迫切需要解决 IP 语义重载的问题.在当前的网络体系结构中,IP 地址既是网络节点的标识(identifier),又是接入位置(locator)的标志,甚至与传输层的套接字紧耦合.如果网络 IP 因为接入位置改变而发生变化,则通信进程将中断,这是体系结构的缺陷.移除方案突出了拓扑无关的 ID 与拓扑相关的 Locator 之间的矛盾,因此,ID/Locator 分离是移除方案在体系结构上改进的重点内容之一.

不同移除方案间的差异主要体现在ID/Locator分离的具体实现方法上.例如,Shim6在网络层和传输层间增加了一个 shim 子层,作为网络标识 ID 的功能,提供协议栈上层需要的接口.原来协议栈的 IP 地址仅作为接入位置 Locator 的功能,传输层端口和 ID 绑定后可作为 session 的标识.接入位置发生改变,或 multihoming 采用不同 IP 地址时,由于网络 ID 保持不变,通信进程(session)不会中断.SIX/ONE 和 NIRA 直接对 IP 地址划分层次,利用 IPv6 地址的后 64 位作为 ID,前缀作为网络的 Locator.GLI 则将 IPv6 地址进一步划分为全局 Locator、本地 Locator 和 ID 这 3 个层次.Name Based Socket 则将 ID 定义到协议栈更上层的位置,将套接字的绑定关系直接转移到域名(fully qualified domain name,简称 FQDN).域名作为 ID,IP 地址作为 Locator,通过新的协议栈实现 ID/Locator 的分离.该方案需要通信双方的主机都支持 IPv6,网络传输则采用 Teredo 6voer4 隧道支持协议过渡,名字和 IP 地址的动态绑定依赖于 DNS 系统.

移除方案改进了体系结构,通过 ID/Locator 分离,解除了拓扑无关的标识和拓扑相关的接入位置之间的耦合关系,有利于采用拓扑相关的 PA 地址提高网络路由的聚合性,能够较为彻底地解决网络路由的可扩展性问题.

移除方案也面临着 3 个难以克服的困难:首先是部署激励的问题.由于移除方案需要在体系结构中嵌入 ID 子层,往往不得不更新现有的协议栈实现.改变协议栈意味着主机端的操作系统以及应用软件的升级.用户和软件开发者往往不得不改变现有协议实现,开销极大.而网络可扩展性问题是网络运营商的主要问题,如果要求端系统和用户为解决网络运营的问题支付成本开销,显然缺乏有效的部署激励机制.从运营商的角度看,ISP 也不一定愿意放弃 PI 地址,因为更改网络 IP 地址(renumbering)可能需要重新配置路由防火墙,更新网络设备的接入控制列表,增加了管理成本;其次,网络地址分配问题.由于网络地址是拓扑相关的,而为了保证用户接入在多个网络接入位置都能采用 PA 地址,需要对每个接入位置分配一个 IP 地址作为网络接入位置定位标识符(locator),这种应用可能导致每个端系统消耗大量地址空间,造成地址空间的浪费.为了保证未来网络的发展规模,multihoming 的站点往往还需要预留多个连续的地址空间.因此,基于地址聚合的“移除”方案地址空间的利用率较低;最后,在移除方案中均采用了 ID/Locator 分离,在网络层路由之前最终需要解析 ID 和 Locator 的映射关系,也不得不在网络中部署映射系统维护这个动态绑定关系,这个映射系统面临分离方案中映射系统同样的一系列问题.

路由聚合的方案可以在地址空间增长的同时保持路由可扩展性,能够满足中(长)期网络规模发展的需要.路由聚合方法的基本前提是拓扑相关的编址.网络拓扑是动态的,而网络标识是静态的,这种矛盾不是简单地通过 ID/locator 分离就可以消弭的.从长远来看,互联网需要新的可扩展路由技术,这一目标激励我们尝试路由体系结构和路由算法.

3.3 改进路由算法与可扩展路由技术

第 3 种提高路由可扩展性的思路着眼于设计新的分布式路由算法,减小路由表尺寸.主要思想是,改变报文转发完全依赖本地路由表信息(或者说节点路由表存储全部路由信息)的路由模式,通过在编址中嵌入网络拓扑信息减小甚至消除节点转发报文的不确定性.此时,报文转发的不确定性不再是等概率的.在我们的路由信息熵理论模型中,等概率转发的不确定性最大,熵有最大值;而非等概率的不确定性,熵值较小.从这个意义上看,在路

© 中国科学院软件研究所 http://www.jos.org.cn

由编址中嵌入网络节点拓扑信息(路由信息)可以降低路由信息熵.或者从另一个角度来解释:路由信息不仅存储在路由表中,还同时嵌入到编址信息中.这两部分信息相结合,最终消除转发过程中的不确定性.因此,分布式路由算法中的“网络拓扑结构”和“适应网络拓扑特性的编址方案”是实现高效可扩展路由的两大关键因素.

研究表明,并不是所有拓扑结构都能有效地实现可扩展路由算法.Krioukov 在文献[28]中综述了紧致路由算法的研究成果,分析了在互联网拓扑上采用紧致路由算法的理论限制,并得出了悲观的结论:连接稀疏的网络和拓扑规则的网络能够比较容易地找到可扩展路由算法,而具有“小世界(small world)”特性的拓扑结构(包括互联网)几乎是所有拓扑中最困难的情形.主要原因是,在动态网络拓扑或拓扑位置无关的网络编址条件下,路由表尺寸和交换路由拓扑信息的报文数量都是不可扩展的(尽管静态拓扑和拓扑相关编址方案可以保证互联网拓扑有可扩展路由算法).

为了在编址中嵌入拓扑信息,可扩展路由算法需要能够准确反映网络拓扑特性的编址方案.具体的编址方案又分为名字相关(name dependent,简称 ND)和名字无关(name independent,简称 NI)两类.名字相关是指,网络节点名字由路由方案指派,并可根据拓扑变化而重命名,基于名字相关编址的可扩展路由方案要求事先获得全网的拓扑信息,这样才能根据拓扑关系进行节点编址.名字无关则允许节点名是任意指定的而且不随拓扑动态而发生改变.这时,通常仍然需要拓扑相关编址方案(名字相关编址方案)的支持,然后将拓扑无关的节点名字映射到拓扑相关的编址.这种机制类似于 ID/Locator 分离和映射,基本的区别在于:在 ID/Locator 分离中,拓扑相关的 Locator 将按拓扑关系聚合,而名字相关的编址没有聚合方面的要求.很明显,实现名字无关的可扩展路由方法更加困难,但却能更好地适应网络拓扑动态.在互联网路由体系结构中,采用名字无关的路由方案更有意义.

为了解决拓扑动态和名字无关编址条件下互联网路由的可扩展性问题,大量理论研究着重分析了 AS 级拓扑的特性,同时借鉴复杂网络可扩展路由算法研究的成果,产生了一系列域间可扩展路由算法.其中,内嵌度量空间的几何路由算法显示出良好的路由性能和一定的应用前景,日益受到关注.通过对复杂网络的研究,人们发现,在“小世界”特性的复杂网络中,节点可以仅通过局部拓扑信息(如邻居节点信息和目的节点信息)实现高效的全局路由发现过程.如果能在网络编址中准确内嵌拓扑特征信息,则在具有这种特性的网络上可以实现高效的几何路由(geometric routing)算法[37]:每个节点仅存储局部的路由信息而同时又能保证较高的路由效率(引入有限的路径延展),因而具有良好的路由可扩展性.

几何路由算法是从早期的地理路由(geographic routing)算法延伸推广而来的,在对 Ad Hoc 网络和某些移动网络的路由算法研究过程中人们发现,可以根据节点的地理坐标进行路由,每个节点向距离目的节点最近的邻居节点转发报文,最终,这种贪婪算法具有很高的可扩展性.但有时贪婪算法也不能保证路由成功,当某节点比它所有的邻居都距目的节点近时,则会产生路由循环.虽然有技术可以避免产生路由循环,但通常会影响路由效率.如果将地理路由的坐标推广为更一般的虚拟坐标,则可以在更抽象的赋范空间实现贪婪策略的几何路由算法.

几何路由算法为了有效地在网络编址中嵌入拓扑信息,通常需要构建一个抽象的度量空间,然后根据节点间的拓扑关系赋予每个网络节点在这个度量空间中的虚拟坐标,这个过程被称为内嵌拓扑度量空间.内嵌拓扑度量空间的技术是拓扑相关编址方法中的关键技术,直接影响到路由的效率.这个拓扑空间可以是一维或多维的,也可以是欧氏空间或非欧空间.当编址采用贪婪内嵌时,可以实现贪婪几何路由算法并保证路由成功.贪婪内嵌定义为:对于无向图 G,如果有映射 $f:V(G)\sim X$ 赋予每个节点 v 在度量空间$(X;d)$上坐标 X,且保证任意不同的两点 $s,t\in V(G)$之间存在 u 与 s 相邻,有 $d(f(u);f(t))<d(f(s);f(t))$.满足这一条件时,要么邻居节点是目的节点,要么总有一个邻居距离目的节点更近.

许多几何路由方案针对网络的不同拓扑特性寻求合适的内嵌度量空间,文献[38,39]揭示了无标度特性的复杂网络中存在隐藏度量空间,并建立了用节点坐标的“空间距离”反映对应节点拓扑特性差异的基本模型.文献[40,41]表明,负曲率双曲空间的几何度量符合无标度(“高成簇性”clustering)网络的拓扑特性,在双曲度量空间可以实现贪婪内嵌,并保证 100%的路由成功率.几何路由算法只需要节点保存直接邻居节点的路由信息,其路由表尺寸仅与节点的度和节点的编址(内嵌坐标)有关,可以保证节点路由表尺寸不超过 $O(\log(n))$,具有良好的

© 中国科学院软件研究所 http://www.jos.org.cn

路由可扩展性.几何路由中,内嵌度量空间过程可以实现分布式的算法,并可以适应拓扑动态,节点加入和退出对拓扑结构的影响是局部的,拓扑改变时不一定需要全局节点重新确定虚拟坐标.从这个意义上看,几何路由是适应动态拓扑特性的可扩展路由算法.

然而,已有的研究表明,通过内嵌拓扑信息编址实现可扩展路由算法仍存在很多局限:首先,网络的拓扑特性需要符合一定要求.不同的拓扑结构在路由发现的效率上存在很大差异,或者说不是所有拓扑都可以实现高效的可扩展路由.经验表明,对于拓扑结构相对规则且稳定的网络,如网格(grid)、树状网和各类超立方体结构(super-cube)等网络,其路由效率较高.因此,基于拓扑信息编址的路由技术在并行高性能计算交换结构中有大量成功的应用.但复杂网络目前仅对某些特殊网络模型验证了路由算法的可扩展性;其次,内嵌拓扑空间的选取必须符合拓扑特性的规律.不同的拓扑特性需要不同的内嵌度量空间编址方案,并且有的路由方案不能保证100%的路由成功率.虽然目前针对复杂网络拓扑建模的研究取得了大量成果,但针对不同的拓扑应当内嵌何种度量空间,以及采用何种策略才能保证路由成功率和较小的路径延展度,仍然没有明确的结论;最后,已有可扩展路由算法仍不能满足域间路由策略表达的需求.因此,这类技术在域间路由的实际应用比较有限.

路由算法革新的思想催生了新的域间路由模式,甚至产生了一些不依赖于现有路由体系结构(“clean slate”)的路由方案.这些方案试图改变节点转发分组时完全依赖本地 IP 路由表的传统寻址转发模式.因此,传统意义上互联网基于 IP 寻址的路由过程被其他路由模式所取代.新的路由体系结构中采用的路由算法不一定需要交换全局拓扑信息就可以满足路由收敛和避免路由循环,在提高路由可扩展性方面,性能更好.这些新路由方案强调有效地利用网络拓扑结构或网络应用本身的特性,是面向未来网络体系结构需求的设计方案.

在面向下一代互联网体系结构的设计领域,GENI,FIND,FIRE 等研究计划分别支持各种创新的路由体系结构.革新思想的出发点是,现有体系结构难以顺利演进,因此不如直接采用新的路由模式,彻底解决当前存在的包括路由可扩展性在内的多种问题.全新的路由体系结构最近日益成为活跃的学术研究领域,其中比较典型的路由方案包括:基于覆盖网(overlay)的路由、基于内嵌拓扑度量空间的几何路由和以内容(content)为中心的数据分发模式.

3.3.1 覆盖网路由模式

覆盖网路由是在IP层之上构建路由机制,IP网络成为下层支持(underlay)网络.这一方法广泛应用于P2P网络.实践证明,可以根据需要,构建需要的逻辑拓扑结构,具有较好的可扩展性.在域间实现覆盖网路由最典型的提案是ROFL(routing on flat labels)[42].

ROFL采用直接基于上层名字空间路由的方案,是应用层覆盖网思想的扩展.ROFL改变了传统的依靠DNS解析网络地址再基于路由表查询建立连接的模式,主机标识(HostID)采用无结构的扁平名字空间(flat label),并在 DHT 结构的分布式系统中建立基于 HostID 的可扩展逻辑路由.这种路由模式允许新方案与 TCP/IP 完全平行发展或建立于 TCP/IP 之上.

ROFL 基于分布式哈希表(DHT)结构构建域内和域间 Overlay 网络,利用 DHT 的结构对扁平名字空间建立逻辑路由路径.该方案采用无层次结构的主机名字空间(HostID),依赖类似于 Chord 的分布式哈希表查询方法建立域内和域间的路由转发机制.如图 2 所示,当主机 a 加入网络时,它的宿主路由器 R_1 建立到负责 id_a 后继 $succ(id_a)$的路由器 R_2 的源路由(source route),并通知负责 id_a 前驱 $predecessor(id_a)$的路由器建立到 R_1 的源路由.最终,这种基于 Chord 的加入算法保证了各主机的 id 形成一个 chord 环,每个 id 所在的路由器都分别建立了到这个 id 前驱和后继的源路由.ROFL 还设计了多种针对覆盖网络的优化技术,以保证路由效率和可靠性.例如:为了减小路径延展,改进了路由表结构以满足 overlay 和 underlay 的拓扑近似性(topology proximity);域边界路由器可选地安装Bloom过滤器,以便快速检测所辖AS内的id;多径underlay连接保证overlay的可靠性等等.ROFL还支持域间策略和 multihoming 以及流量工程的路由需求.

ROFL 的覆盖网路由避免了网络地址重新分配(renumbering)和 ID/Locator 动态映射的问题.同时,对主机接入控制直接建立在对名字的控制上而不依靠网络接入位置.但基于逻辑覆盖网络路由的模式需要 underlay 网络的支持,ROFL 的 underlay 采用源路由(source routing)模式,转发分组中带有源路由信息,underlay 节点间采用

© 中国科学院软件研究所 http://www.jos.org.cn

类似链路 OSPF 模式的路由协议,维护一定范围内不同层次(域内/域间)的 underlay 网络拓扑.

覆盖网路由存在的主要问题是,overlay 和 underlay 之间的拓扑一致性不一定总能得到有效保证.另外,underlay 的路由效率对 overlay 有直接和必然的影响,由此而产生的路径延展和延时都没有确定的上界.在 P2P 网络中,许多优化技术由于域间策略的原因也无法实现.这些都限制了覆盖网路由在域间路由模式中的应用效率.

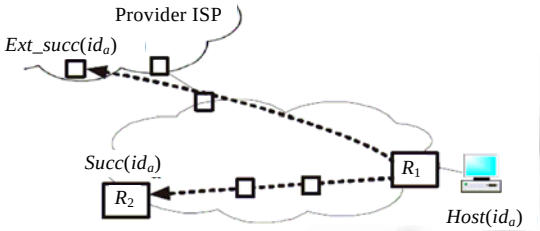

Fig.2 ROFL DHT routing

图 2 ROFL DHT 路由

3.3.2 基于内嵌拓扑度量空间的几何路由

可扩展路由的理论研究表明,几何路由算法是一类有应用前景的可扩展路由算法,能够适应动态拓扑并实现名字无关编址.已有理论证明,任何名字无关路径延展度为 1 的路由算法,必须全局交换拓扑信息,且路由表规模为$\Omega(n)$.为了提高路由可扩展性,必须保留一部分路由不确定性,结果必然引入路径延展(stretch).近年来,大量可扩展路由算法试图探寻在路由信息不完备情况下(不同路由表尺寸)各种拓扑具有的路径延展度的下限,并取得了一定的成果.事实证明,在某些特殊拓扑上,即使没有确定的路由信息,仍可以依据特殊的路由算法保证以较高的概率正确转发(较低的平均路径延展).通常,在具有无标度(scale free)特性的复杂网络上可以实现高效的路径发现过程.于是,针对互联网拓扑特性的可扩展路由研究,尝试用路径搜索取代传统的寻址转发模式,即节点不再存储全局的路由信息,而是基于本地局部路由信息和目的节点地址进行转发推断,最终实现近似最优的路由转发功能.此时,路由表尺寸仅与直连邻居规模(节点的度)相当,因此具有良好的路由可扩展性.

最近的研究发现,互联网 AS 级拓扑符合无标度网络的基本特征,度分布服从幂律$\rho(k)=k^{-\gamma}$,幂指数 $2.1<\gamma<2.3$,并且具有较高的成簇性(clustering).这种拓扑结构表现出较强的自相似性,十分适合嵌入负曲率的几何度量空间.许多研究成果表明,内嵌双曲度量空间的几何路由[40,41]是应用在域间拓扑结构上较理想的贪婪路由算法.

基于内嵌度量空间的几何路由算法是名字无关的可扩展路由方法,能够应对网络拓扑的动态性,在服从幂律的拓扑模型上引入的路径延展度较小(平均 $stretch<3$).该方法具有良好的可扩展性,因为路由表尺寸仅与节点度和编址方案相关.该方法的限制在于拓扑特性比较苛刻,当幂指数$\gamma>3$ 时,路径延展随即增大;并且,在不满足贪婪内嵌的条件时,不能保证贪婪路由算法的成功率.因此,尽管几何路由算法显示出优越的可扩展性,但在互联网未来拓扑演进规律不确定的情况下,该方法对互联网路由体系结构的影响仍然不够明确.

3.3.3 面向内容的数据分发模式

目前,互联网路由功能是面向寻址空间的分布式定位服务.路由转发所涉及的拓扑信息交互过程以通信节点的拓扑位置为中心,但许多网络应用更关心数据内容本身而并非存储数据的节点.一个大胆的革新想法因此应运而生,即网络应用面向内容(content)而非产生/承载内容的实体.这样一来,网络通信不再以寻址为基础,获得或交换节点的拓扑位置信息也不一定是网络交换数据的先决条件.网络体系结构从传统的分布式寻址转发路由模式演化为面向内容的分布式存储和数据分发系统,网络层以寻址为中心的路由模式转变为以内容为中心的数据分发模式.

这种新模式打破了节点和数据间一一映射的绑定关系.网络应用获得数据不是首先将数据映射为节点位置(DNS 功能),然后再寻址路由到定位节点的位置(路由功能).在面向内容的数据分发模式下,数据(或产生数据的实体)可以在网络中有多个拷贝,因此往往有多个接入位置.这种模式解除了数据传输意义上的路由转发定位

© 中国科学院软件研究所 http://www.jos.org.cn

与内容本身的一对一绑定关系.另外,内容/数据本身在网络中可以有一定的逻辑层次结构,网络应用根据内容或数据的组织层次结构向邻居节点发送请求数据的消息,这个数据发现过程以类似洪泛的方式向周边网络节点扩散.因此,网络寻址转发功能被弱化,网络设备面向内容(数据)本身进行分布式查询和交换,而不再关心数据的位置和路由转发过程,因此可以不再需要全局路由信息,节点仅需要局部邻居的拓扑连接信息即可,路由表规模取决于内容名字的聚合程度和拓扑连接的度.较典型的提案包括 DONA(data oriented network architecture)[43] 和 CCN(content centric network)[44].

这类新方案还处于研究初期,目前,其性能评价还缺乏实际应用的验证.但是,通信模式的转变对互联网体系结构的影响是革命性的.新的模式反映了当前互联网应用的特点和新需求,该思想集中体现了现代互联网中数据密集型应用占主导地位的特点和大规模自组织网络对路由可扩展性的需求.该思想将互联网体系结构设计带入一个全新的领域,有许多未被认知的规律有待未来的研究和进一步深入探索.

上述 3 类解决域间路由可扩展性问题的基本思路分别针对路由信息熵理论模型分析中的 3 个关键影响因素(即地址空间、路由聚合程度和网络拓扑特性),指明了解决路由可扩展性问题的不同方向.

4 域间可扩展路由设计的挑战性问题

根据前文基本理论模型的分析和对近年来路由方案研究的分析评价,互联网域间可扩展路由设计具有如下几个挑战性问题:

- 如何适应网络拓扑特性及拓扑演进规律

已有的研究表明,互联网当前的拓扑特性不利于实施层次化路由[25],也不一定保证有高效的可扩展路由算法[28].在网络规模急剧扩张的同时,其拓扑演进规律尚不明确.目前比较清楚的是对理想的域间可扩展路由算法的基本要求:名字无关的方案;最大路径延展较小(近似最优);随网络规模对数级增长的路由表尺寸和拓扑动态更新开销.能否在具有无标度特征的复杂网络上合理地构建路由层次或较理想地实现可扩展路由算法,是具有挑战性的问题之一.

- 如何弥补现有体系结构缺陷

在现有路由体系结构中,网络层的语义重载问题直接关系到拓扑相关编址和拓扑无关标识之间的矛盾.ID 和 Locator 分离,并按拓扑关系编址,成为路由聚合的前提.解决这个矛盾需要考虑当前路由机制的向后兼容问题.目前,除了地址翻译和隧道封装外,还缺乏其他成熟的应用技术.在 ID/Locator 分离方案中,映射系统也是必须面对的关键问题,其效率和可扩展性在大规模部署前尚无法有效验证.如何在完善路由体系结构的同时提高路由聚合程度并最终解决路由可扩展性问题,是另一个具有挑战性的问题

- 如何支持路由策略和管理模式

由于互联网自治系统采用相对独立的管理运营模式,因此缺乏网络统筹规划和路由协调机制.各 AS 通过路由协议独立表达路由策略,这些路由策略的原则首先必须符合商业关系和本地管理运营利益的最大化,这个原则往往与全局要求路由聚合的需求相矛盾.基本现实是:最长前缀匹配技术加剧了地址前缀的碎片化,是导致路由可扩展问题的主要原因之一.如何设计既能满足 AS 策略灵活表达又不致影响路由聚合性的路由模式,是又一个具有挑战性的问题.

- 如何同时满足其他体系结构设计的需求

设计新的路由体系结构,可扩展性只是其中重要的特性之一,但网络设计还需要同时满足安全可信性、泛在性、移动性、可演进性等其他基本特性.如何避免可扩展性与其他设计原则相矛盾,是一个基本的挑战性问题.

上述 4 个问题难度逐次增加.新一代互联网体系结构设计需要研究新的指导理论,统筹兼顾上述设计需求,该领域是今后互联网发展历程中需要进一步展开的重点范畴之一.

© 中国科学院软件研究所 http://www.jos.org.cn

5 展 望

下一代互联网研究机遇和挑战并存,路由可扩展性是互联网发展的最基本要求之一,这一问题是互联网设计之初无法预计的.随着网络规模的持续扩张,网络应用的不断丰富,这个问题将越来越突出,体系结构上的矛盾也会日益尖锐.

随着IPv4地址的即将耗尽,IPv6即将大规模部署,我们有理由相信,未来的网络规模将超出现有路由系统的可扩展能力.虽然当前 IPv6 的部署和应用规模尚十分有限,从 BGP 观测点获得的数据表明,IPv6 路由表项仅几千条.但是IPv6的地址空间十分巨大,因此对未来路由系统的可扩展能力提出了更高的要求.如果不采用合理的路由机制(如限制 BGP 宣告前缀的最小长度),则路由系统的可扩展性堪忧.IPv6 路由系统是否有不同的扩展特性,取决于其基本路由模式和路由机制的限制,这个问题有待将来进一步加以研究.

如果对下一代互联网作最大胆的设想:未来互联网的路由规模决定于网络应用的泛在性,特别是物联网应用的极大需求将极大地推进网络规模的扩张.在这样的应用背景下,路由系统要应对更大的地址空间,必须有高性能的路由协议满足高带宽的路由转发功能的需求.路由系统的可扩展性是网络体系结构的基本要求之一.为了满足这一要求,有必要合理构建路由层次,可以考虑适当增加路径延展度,牺牲少量网络的平均跳数.新的体系结构下也有可能部分地采用面向内容本身的数据分发新机制,分化路由转发的功能属性.从某种意义上说,网络的即时通信功能和非实时的共享数据分发功能可能需要在路由层面上分离,以减轻大规模路由转发的压力.如果体系结构的功能异化最终导致互联网路由功能的分化,则路由可扩展性问题本身的涵义将发生变迁.总之,路由系统的可扩展性问题仍然是一个需要深入研究的开放问题.

致谢 本文受益于网络体系结构研究组的胡虹雨博士及研究组其他博士生的研究和讨论,在此一并致谢.

References:

[1] Doria A, Davies E, Kastenholz F. A set of possible requirements for a future routing architecture. RFC 5772, 2010.

[2] Davies E, Doria A. Analysis of inter-domain routing requirements and history. RFC 5773, 2010.

[3] Meyer D, Zhang L, Fall K. Report from the IAB workshop on routing and addressing. RFC 4984, 2007.

[4] Krioukov D, Fall K, Yang X. Compact routing on Internet-like graphs. In: Proc. of the IEEE INFOCOM 2004. Piscataway: IEEE, 2004. 209−219.

[5] Li T, ed. Design goals for scalable internet routing. Internet-Draft, 2007.

[6] Zhang LM. Complex network and compact routing. ZTE Communications, 2009,15(6):5−8 (in Chinese with English abstract).

[7] Cui Y. IETF highly concerns routing scalability problems. China Educational Network, 2007 (in Chinese with English abstract). http://www.cnki.com.cn/Article/CJFDTOTAL-JYWL200704022.htm

[8] Zhan FB, Xu MW, Wu JP, Survey on host identity protocol (HIP). Journal of Chinese Computer Systems, 2007,28(2):224−228 (in Chinese with English abstract).

[9] Yu SP. Study on the locator/identifier separation. Computer Technology and Development, 2007,19(7):95−97 (in Chinese with English abstract).

[10] Xu XH, Guo DY, Gao XS, Cao W, Li HJ. Discussion on Internet scalability problems. Telecommunications Network Technology, 2009,4:6−11 (in Chinese with English abstract).

[11] Li JR. Scalability of Internet. Science and Technology Information, 2009,21:65−66 (in Chinese with English abstract).

[12] Shannon CE. A mathematical theory of communication. Bell System Technical Journal, 1948,27:379−423, 623−656.

[13] Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the Internet topology. In: Proc. of the Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communication. New York: ACM, 1999. 251−262.

[14] Massey D, Wang L, Zhang B, Zhang L. A scalable routing system design for future Internet. In: Proc. of the ACM SIGCOMM Workshop on IPv6. Kyoto: ACM, 2007. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.120.218&rep=rep1&type=pdf

[15] Subramanian L, Caesar M, Cheng TE, Handley M, Mao M, Shenker S, Stoica I. HLP: A next generation inter-domain routing protocol. Computer Communication Review, 2005,35(4):13−24.

© 中国科学院软件研究所 http://www.jos.org.cn

[16] Farinacci D, Fuller V, Meyer D, Lewis D. Locator/ID separation protocol (LISP). Internet-Draft, 2009. http://tools.ietf.org/html/draft-ietf-lisp-08.txt

[17] Whittle R. Ivip (Internet vastly improved plumbing) architecture. Internet-Draft, 2010. http://tools.ietf.org/html/draft-whittle-ivip-arch-04

[18] Jen D, Meisel M, Massey D, Wang L, Zhang B, Zhang LX. APT: A practical transit mapping service. Internet-Draft, 2007. http://tools.ietf.org/html/draft-jen-apt-01

[19] Adan JJ. Tunneled inter-domain routing (TIDR). Internet-Draft, 2006. http://tools.ietf.org/html/draft-adan-idr-tidr-00

[20] Templin F, ed. The IPvLX architecture. Internet-Draft, 2007. http://tools.ietf.org/id/draft-templin-ipvlx-08.txt

[21] Zhang B, Zhang L, Wang L. Evolution towards global routing scalability. Internet-Draft, 2009. http://tools.ietf.org/html/draft-zhang-evolution-02

[22] Zhang B, Wang L, Zhao X, Liu Y, Zhang L. FIB aggregation. Internet-Draft, 2009. http://tools.ietf.org/html/draft-zhang-fibaggregation-02

[23] Ballani H, Francis P, Cao T, Wang J. Making routers last longer with ViAggre. In: Proc. of the 6th USENIX Symp. on Networked Systems Design and Implementation. Berkeley: USENIX Association, 2009. 453−466.

[24] Zhang XY, Francis P, Wang J, Yoshida K. Scaling IP routing with the core router-integrated overlay. In: Proc. of the 2006 IEEE Int'l Conf. on Network Protocols. Washington: IEEE Computer Society, 2006. 147−156.

[25] Kleinrock L, Kamoun F. Hierarchical routing for large networks: Performance evaluation and optimization. Computer Networks, 1997,1(3):155−174.

[26] Castineyra I, Chiappa N, Steenstrup M. The nimrod routing architecture. RFC 1992, 1996. http://tools.ietf.org/html/rfc1992

[27] Kastenholz F. ISLAY: A new routing and addressing architecture. Internet-Draft, 2002. http://tools.ietf.org/all-ids/draft-irtf-routing-islay-00.txt

[28] Krioukov D, Claffy K, Fall K, Brady A. On compact routing for the Internet. Computer Communication Review, 2007,37(3):41−52.

[29] Nordmark E, Bagnulo M. Shim6: Level 3 multihoming shim protocol for IPv6. RFC 5533, 2009. http://tools.ietf.org/html/rfc5533

[30] Vogt C. Six/One router: A scalable and backwards compatible solution for provider-independent addressing. In: Proc. of the 3rd Int'l Workshop on Mobility in the Evolving Internet Architecture. Seattle: ACM, 2008. 13−18.

[31] Yang X, Clark D, Berger AW. NIRA: A new inter-domain routing architecture. IEEE/ACM Trans. on Networking, 2007,15(4): 775−788. [doi: 10.1109/TNET.2007.893888]

[32] Atkinson R. ILNP concept of operations. Internet-Draft, 2010. http://ilnp.cs.st-andrews.ac.uk/docs/id/draft-rja-ilnp-intro-03.txt

[33] Menth M, Hartmann M, Klein D. Global locator, local locator, and identifier split (GLI-Split). Report No.470, University of Wurzburg Institute of Computer Science Research Report Series. Wurzburg: University of Wurzburg, 2010. http://www3.informatik.uni-wuerzburg.de/~menth/Publications/papers/Menth-GLI-Split.pdf

[34] Moskowitz R, Nikander P. Host identity protocol (HIP) architecture. RFC 4423, http://tools.ietf.org/html/rfc4423

[35] Frejborg P. Hierarchical IPv4 framework. Internet-Draft, http://tools.ietf.org/html/draft-frejborg-hipv4-08

[36] Feldmann A, Cittadini L, Mühlbauer W, Bush R, Maennel O. HAIR: Hierarchical architecture for Internet routing. In: Proc. of the 2009 Workshop on Re-Architecting the Internet. New York: ACM, 2009. 43−48.

[37] Kuhn F, Wattenhofer R, Zhang Y, Zollinger A. Geometric ad-hoc routing: Of theory and practice. In: Proc. of the 22nd Annual Symp. on Principles of Distributed Computing. New York: ACM, 2003. 63−72.

[38] Serrano MA, Krioukov D, Boguna M. Self-Similarity of complex networks and hidden metric spaces. Physical Review Letters, 2008,100(7):078701-1. [doi: 10.1103/PhysRevLett.100.078701]

[39] Boguna M, Krioukov D, Claffy K. Navigability of complex networks. Nature Physics, 2009,5:74−80. [doi: 10.1038/nphys1130]

[40] Shavitt Y, Tankel T. Hyperbolic embedding of Internet graph for distance estimation and overlay construction. IEEE/ACM Trans. on Networking, 2008,16(1):25−36. [doi: 10.1109/TNET.2007.899021]

[41] Kleinberg R. Geographic routing using hyperbolic space. In: Proc. of the 26th IEEE Int'l Conf. on Computer Communications (IEEE INFOCOM 2007). Piscataway: IEEE, 2007. 1902−1909. [doi: 10.1109/INFCOM.2007.221]

[42] Caesar M, Condie T, Kannan J, Lakshminarayanan K, Stoica I, Shenker S. ROFL: Routing on flat labels. In: Proc. of the SIGCOMM 2006. New York: ACM, 2006. 363−374.

© 中国科学院软件研究所 http://www.jos.org.cn

[43] Koponen T, Chawla M, Chun B, Ermolinskiy A, Kim KH, Shenker S, Stoica I. A data-oriented (and beyond) network architecture. In: Proc. of the SIGCOMM 2007. New York: ACM, 2007. 181−192.

[44] Jacobson V, Smetters D, Thornton J, Plass M, Briggs N, Braynard R. Networking named content. In: Proc. of the CoNEXT 2009. New York: ACM, 2009. 1−12.

附中文参考文献:

[6] 张连明.复杂网络与可扩展路由.中兴通信技术,2009,15(6):5−8.

[7] 崔勇.IETF 高度关注路由可扩展性问题.中国教育网络,2007. http://www.cnki.com.cn/Article/CJFDTOTAL-JYWL200704022.htm

[8] 昝风彪,徐明伟,吴建平.主机标识协议(HIP)研究综述.小型微型计算机系统,2007,28(2):224−228.

[9] 于世鹏.标识符和定位符分离方案研究.计算机技术与发展,2007,19(7):95−97.

[10] 徐小虎,郭大勇,高雪松,曹玮,李贺军.一种解决路由可扩展问题的网络新架构:虚拟聚合(VA).电信网技术,2009,4:6−11.

[11] 李继荣.Internet 的可扩展性.科技信息,2009,21:65−66.

张威(1972－),男,湖北武汉人,博士生,讲师,主要研究领域为计算机网络体系结构,Internet 域间路由.

吴建平(1953－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算机网络体系结构,网络协议工程学,下一代互联网.

毕军(1972－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算机网络体系结构,网络安全,协议工程学.

© 中国科学院软件研究所 http://www.jos.org.cn