Internet 拓扑建模综述*

张 宇+, 张宏莉, 方滨兴

(哈尔滨工业大学 计算机网络与信息安全技术研究中心,黑龙江 哈尔滨 150001)

A Survey on Internet Topology Modeling

ZHANG Yu+, ZHANG Hong-Li, FANG Bin-Xing

(Research Center of Computer Network and Information Security Technology, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: Phn: +86-451-86418272, E-mail: zhangyu@pact518.hit.edu.cn, http://www.hit.edu.cn

Received 2003-09-15; Accepted 2004-03-31

Zhang Y, Zhang HL, Fang BX. A survey on Internet topology modeling. *Journal of Software*, 2004,15(8): 1220~1226.

http://www.jos.org.cn/1000-9825/15/1220.htm

Abstract: As the basis of Internet development and exploitation on higher levels, the Internet topology modeling starts from the random model to the hierarchical model. Then it developed to a more realistic one, scale-free network model. Many characteristics of topology are analyzed with the corresponding metrics, including power law. Moreover, the related work on the current topology models, topology generation algorithms, and topology generators is fully presented. Finally, the new problems and challenges which arise from current research are discussed and some suggestions for future research work are put forward.

Key words: Internet topology modeling; network modeling; topology generator; topology metric; Internet measurement and analysis; power law; scale-free network

摘 要: Internet 拓扑建模是在更高层次上开发、利用 Internet 的基础.Internet 拓扑模型研究经历了从随机型到层次型,再到无尺度(scale-free)网络的过程.对包括幂率(power law)在内的多种 Internet 拓扑特征及其相应度量进行了分析,对现有的拓扑模型、拓扑生成算法以及拓扑生成器进行了全面的综述.最后论述了目前研究中遇到新的问题与挑战,并对今后技术路线进行了总结.

关键词: Internet 拓扑建模;网络建模;拓扑产生器;拓扑度量;Internet 测量与分析;幂率;无尺度网络

中图法分类号: TP393 **文献标识码**: A

Internet 作为当今人类社会信息化的标志,其规模正以指数速度高速增长.如今 Internet 的“面貌”已与其原型 ARPANET 大相径庭,依其高度的复杂性,可以将其看作一个由计算机构成的“生态系统”.虽然 Internet 是人类亲手建造的,但却没有人能说出这个庞然大物看上去到底是个什么样子,运作得如何.Internet 拓扑建模研究就

* Supported by the National Natural Science Foundation of China under Grant No.60203021 (国家自然科学基金)

作者简介: **张宇**(1979－),男,河北乐亭人,硕士生,主要研究领域为网络信息安全,并行处理;**张宏莉**(1973－),女,博士,副教授,主要研究领域为网络信息安全,并行处理;**方滨兴**(1960－),男,博士,教授,博士生导师,主要研究领域为网络信息安全,并行处理.

© 中国科学院软件研究所 http://www.jos.org.cn

是探求在这个看似混乱的网络之中蕴含着哪些还不为我们所知的规律.发现 Internet 拓扑的内在机制是认识 Internet 的必然过程,是在更高层次上开发利用 Internet 的基础.然而,Internet 与生俱来的异构性、动态性、发展的非集中性以及如今庞大的规模都给拓扑建模带来巨大挑战[1].Internet 拓扑建模至今仍然是一个开放性问题,在计算机网络研究中占有重要地位.

Internet 拓扑作为 Internet 这个自组织系统的“骨骼”,与流量、协议共同构成模拟 Internet 的 3 个组成部分,即在拓扑网络中节点间执行协议,形成流量.Internet 拓扑模型是建立 Internet 系统模型的基础,由此而体现的拓扑建模意义也可以说就是 Internet 建模的意义,即作为一种工具,人们用其来对 Internet 进行分析、预报、决策或控制.Internet 模型中的拓扑部分刻画的是 Internet 在宏观上的特征,反映一种总体趋势,所以其应用也都是在大尺度上展开的.对 Internet 拓扑模型的需求主要来自以下几个方面:(1) 许多新应用或实验不适合直接应用于 Internet,其中一些具有危害性,如蠕虫病毒在大规模网络上的传播模拟[2];(2) 对于一些依赖于网络拓扑的协议(如多播协议[3]),在其研发阶段,当前 Internet 拓扑只能提供一份测试样本,无法对协议进行全面评估,需要提供多个模拟拓扑环境来进行实验;(3) 从国家安全角度考虑,需要在线控制网络行为,如美国国防高级研究计划局(DARPA)的 NMS(network modeling and simulation)项目[4].

Internet 拓扑建模是一项复杂的工作,涉及网络测量、图论、算法设计、统计学、数据挖掘、可视化以及数学建模等多个研究领域.Internet 拓扑建模研究内容可归结为 3 个问题:(1) 如何获得一份完整而准确的 Internet 拓扑数据;(2) 如何对 Internet 网络拓扑特征进行描述;(3) 如何构造一幅类似于 Internet 的拓扑图.这 3 个问题分别对应于 3 个研究方向,即拓扑结构测量、拓扑特征发现、拓扑生成器开发.其中拓扑结构测量是拓扑建模的基础;拓扑特征发现是拓扑建模的核心;拓扑生成器就是拓扑模型的软件实现.拓扑结构测量属于网络测量范畴,详细内容参见文献[5].本文将论述后两个方向.

1 拓扑建模研究历程

Internet 拓扑研究经历了从经验假设到客观分析,从单纯的计算机网络研究到复杂系统特征化研究的过程.在研究早期,由于缺乏真实测量数据支持,拓扑模型都是研究人员基于经验假设建立的.最早的 Internet 拓扑模型是 1988 年 Waxman 提出的 Waxman 模型[6].这是一种随机模型,沿用了很多年.直到 1996 年,Doar 提出了 Tiers(等级)模型[7].该模型刻画了 Internet 所具有的层次特征.1997 年,Zegura 等人[8]提出了另一种层次模型——Transit-Stub 模型.此时,从 1995 年开始的大规模 Internet 拓扑测量工作已经逐步展开,采集到了大量拓扑数据.这些数据对科研机构都是免费开放的,极大地推动了拓扑研究的发展.Internet 拓扑研究也进入了一个成果不断累加的阶段,发现多于改进,新旧成果共同描绘了 Internet 拓扑图景.在所有研究成果中,最重要的是,1999 年 Faloutsos 等人发现 Internet 拓扑结构中存在幂律(power-law)[9].幂律的发现将 Internet 拓扑与一些生物学、社会学中的复杂网络联系起来,使其成为“无尺度(scale-free)”网络的一个实例,在 Internet 拓扑研究与系统学研究之间架起了一座桥梁[10].2000 年以来,研究人员开发了许多遵循幂律的拓扑生成算法[11~16]以及拓扑生成器[17~19],为 Internet 模拟提供了有利的支持.不过,现有成果距离前面提到的 3 个问题的答案还有一定距离,而且新的发现[20,21]也对现有成果提出了挑战.

目前,对拓扑模型、生成算法和拓扑生成器的含义的界定比较模糊,研究过程中也常将前两者的软件实现与专门的拓扑生成器拿到一起来比较,或者将生成算法也称为拓扑模型.造成这一情况的原因是,一些生成算法涉及到了 Internet 的内在机制,如 ESF[14];而一些拓扑生成器又包含了一些 Internet 细节特征,如 nem[19].实际上,这三者是有本质区别的:拓扑模型是用数学语言对 Internet 宏观拓扑特征进行刻画,并不一定要提出如何构造这一特征;生成算法是描述如何产生一幅具有某个模型特征的拓扑图,不需要软件化;拓扑生成器则需要根据生成算法来产生一幅类似实际 Internet 的拓扑图,需要从应用角度加以考虑.

Internet 拓扑通常进行形式化为无向图 $G,G=(V,E)$,其中 $V=\{v|v$ 是 G 中的一个节点$\}$,$E=\{(u,v)|u,v\in V$ 且 u,v 相邻接$\}$,G 数学表示为邻接矩阵 M.设 d_v 表示节点 v 的出度,$d_v=|\{u|(u,v)\in E\}|$.设 f_d 表示节点出度频率,$f_d=|\{v|v\in V$ 且 $d_v=d\}|$.

下面我们就拓扑模型、生成算法和拓扑生成器逐一展开讨论,探讨其研究进展与关键技术.

© 中国科学院软件研究所 http://www.jos.org.cn

2 拓扑模型

Internet 拓扑模型可分为两类:一类是描述 Internet 拓扑特征,包括 Waxman 模型,Tiers,Transit-Stub 和幂律;另一类是描述拓扑特征形成的机理,包括 Barabási 与 Albert 提出的 BA 和 ESF 以及一种改进模型 GLP.由于后一类模型对 Internet 幂律形成机理的描述还不成熟,更多的是作为一种产生幂律的图生成算法,所以将在第 3 节进行论述.

对于第 1 类模型来说,Internet 拓扑特征的发现,实际上就是刻画该特征的度量(metric)的发现.一个属于第一类的拓扑模型就是包含若干已存在的或新发现的度量,然后根据实际的 Internet 拓扑数据求出这些度量的值.因此,对这类模型进行评价需要从两方面入手:一方面,对它所采用的拓扑数据进行评价;另一方面,对其度量进行评价.在所有已经发现的 Internet 拓扑度量中,最为基本的就是节点出度频率 f_d.其分布是判断一幅拓扑图是否与 Internet 拓扑相类似的最重要的依据.在研究早期,研究人员或者认为 Internet 节点出度是完全随机的(如 Waxman 模型[6]),或者认为节点出度是正规的(如 Tiers 模型[7]).而幂律的发现证明了 Internet 拓扑结构介于两者之间,呈幂率分布[9].根据对出度频率分布的刻画,可将 Internet 拓扑模型分为以下 3 类:

(1) 随机型,即认为 Internet 拓扑图处于一个完全无序的状态,在大尺度上是均一的.Waxman 模型[6]是一种类似于 Erdös-Rényi 模型的随机模型,出度频率呈泊松分布.这个模型有两个版本:RG1,先将节点随机布置在直角坐标网格中,节点间的距离就是其欧几里德距离;RG2,依据(0,L)均匀随机分布为节点对指定距离.两个版本中,节点间相接的概率 $P(u,v)$与其距离相关,服从泊松分布,距离越近,概率越大.

$$P(u,v)=\beta\exp\frac{-d(u,v)}{L\alpha} \tag{1}$$

其中,$d(u,v)$表示节点 u 与 v 之间的距离,L 为节点间最长距离,α 与 β 取值范围是(0,1).

(2) 层次型,来自对 Internet 结构所具有的层次特征的认识,在同一层上的节点出度接近,不同层间节点出度差别很大.对同一层上的节点布置借用 Waxman 模型方法.Tiers(等级)模型[7]将 Internet 划分为 LAN(局域网),MAN(城域网)和 WAN(广域网)3 个层次.在该模型中,WAN 只有一个,通过指定 LAN 和 MAN 数量以及各自内部所包含节点的数量来构造拓扑图.Transit-Stub 模型[8]将 AS 域划分为 Transit 类和 Stub 类.在该模型中,Transit 节点彼此互联构成一个节点群,一个或多个 Transit 节点群构成拓扑图的核心,而 Stub 节点分布在 Transit 节点群四周与 Transit 节点相连.Transit-Stub 是 GT-ITM(georgia tech Internetwork topology models)软件包的一部分,有时 GT-ITM 也就是指 Transit-Stub 模型.

(3) 幂率型.1999 年,Faloutsos 等人[9]对 NLANR(National Lab for Applied Network Research)在 1997~1998 年间的 3 份 BGP 数据以及 1995 年的一份 traceroute 测量数据进行分析,发现 Internet 拓扑中存在着 4 条幂律.幂律是指形如 $Y\propto X^C$ 的方程,对于两个变量 X 和 Y,存在一个常数 C,使得 Y 与 X 的 C 次幂成比例.

有两个声明:(1) 节点 v 的等级为 r_v,v 是在按出度降序排列序列中的索引值;(2) 邻接矩阵特征值按降序排列,第 i 个特征值为λ_i.

幂律 1(等级指数 R):节点出度 d_v 与该节点等级 r_v 的 R 次幂成比例.

幂律 2(出度指数 O):出度频率 f_d 与该出度 d 的 O 次幂成比例.

近似幂律(hop-plot 指数 H):h 跳内节点对(pairs of nodes)的数量与 h 的 H 次幂成比例.

幂律 3(特征指数 E):特征值λ_i 与其次序 i 的 E 次幂成比例.

幂律 1 说明在 Internet 中既不存在 Waxman 模型中那样的“平等”,也不像 Tiers 和 Transit-Stub 模型那样“等级森严”,而是具有一种松散的层次性.幂律 1 和幂律 2 反映出这些系统具有高度非均一性,即少数节点具有很高的出度,而大量节点具有较小的出度,直接否定了 Waxman 模型中节点分布规律.近似幂律中的 hop-plot 指数 H 可以用来对拓扑图进行分类,例如,对于环,H=1;对于二维网格,H=2.对于 Internet,在 AS 级上,$H\approx4.7$[9];在路由器级上,$H\approx2.8$[9].幂律 3 可以用来进一步对相似的同类图形进行区分.幂律的发现也给出一个启示:Internet 中某个属性值的平均值不能准确地刻画这一属性,应该尝试用某个指数来进行刻画.

对幂律的概念化形象描述是 2001 年 Tauro 等人[22]针对幂律 1 和幂律 2 提出的一种概念模型.这个模型将

© 中国科学院软件研究所 http://www.jos.org.cn

AS 级拓扑图看作一只水母:水母的帽子代表 Internet 的核心,随着与核心的远离,节点重要性逐渐下降,中间的腿比两边的长表示出度为 1 的节点具有聚集性;网络对随机故障是鲁棒的,而若在核心出了故障却是灾难性的.幂律的发现激发了研究人员在 Internet 拓扑结构中寻找更多的规律,也引起了学术界的广泛关注.2001 年,Magoni 和 Pansiot[23]对 NLANR 在 1997~2000 年间的 BGP 数据进行更细致的分析,提出了 Internet 拓扑发展的 5 条经验规律以及 4 条关于结点间最短路径和树结构的幂律.2003 年,Chalmers 等人[3]发现,在多播树的某些特征中也存在类似幂律的分布.

除了上面提到的度量之外,还存在许多其他拓扑度量来描述 Internet 拓扑特征,包括:

1998 年,Watts 和 Strogatz 在他们的“small-world”理论[24]中引入了两个度量:(1) 平均路径长度,也称为特征路径长度,描述图的疏密程度,等于图中每对节点之间最短路径距离的平均值;(2) 聚类系数(clustering coefficient),用来描绘图中出现的小集团.设节点 v 有 d_v 个邻居,那么邻居之间最多有 $M=d_v(d_v-1)/2$ 条边,C_v 等于邻居间实际边数除以 M 的商,再将每个节点的 C 取平均就得到聚类系数.

2002 年,Tangmunarunkit 等人[25]在对层次模型与幂律模型进行比较的过程中引入了 3 个度量.(1) 膨胀(expansion),描述扩散速度.对每个节点计算其在有限跳内可达的节点数量,然后取平均值,再用节点总数来规格化.(2) 恢复(resilience),描述备用路径的存在性.将一个连通图划分为近似相等的两个连通图的最小割集大小,反映图对于连接故障的鲁棒性.(3) 扭曲(distortion),描述一个图像一棵树的程度.对图 G 的一棵生成树 T,G 中相邻两点在 T 上距离的平均值就是 D_T 值,这表示如果把路径限定在 T 上,G 中相邻的两点需要走多少额外的跳数,图 G 的扭曲值 D 就是所有生成树的 D_T 值中最小的.

另外,还包括:(1) 平均偏心度(average eccentricity):每个节点的偏心度是它到其他节点距离中的最大值,平均偏心度就是所有节点偏心度的平均值,反映了图的直径.(2) 最大团尺寸(maximum clique size):节点两两相连的图称为团,Internet 拓扑的最大团就是 Internet 的核心.(3) 正规拉普拉斯谱(normalized Laplacian spectrum). 2003 年,Gkantsidis 等人[26]采用谱过滤技术对 AS 级拓扑邻接矩阵进行正规拉普拉斯谱分析,反映出拓扑结构中节点群及其等级的高低.

3 拓扑生成算法

除了 Waxman,Tiers 和 Transit-Stub 模型中的 3 种拓扑图生成算法之外,目前共有 5 种遵循幂律的拓扑图生成算法以及 1 种图取样(map sampling)算法.这 5 种幂律拓扑生成算法都是针对幂律 2 的,而其他拓扑度量都是在拓扑生成器中进行优化的.5 种幂律算法可以分为两类:一类是直接按照幂律来为节点分配出度,包括 PLOD 和 PLGR(Model A);另一类则通过模拟 Internet 进化过程来满足幂律,包括 BA,ESF 和 GPL.

PLOD(power-law outdegree)[11]是 Palmer 等人于 2000 年提出来的.算法首先确定图中节点数量,根据幂律中的指数分布为每个节点分配一个出度贷款(credit).然后循环从图中随机选取一对节点,如果这对节点不同且尚未连接,并且各自的出度贷款还有剩余,则将它们相连,各自的出度贷款减 1,直到再也找不到满足上述条件的节点对为止.

PLRG(power-law random graphs)[12]是 2000 年由 Aiello 等人提出来的,也被称为 Model A.算法先按照幂律为每个节点分配出度,然后构造一个包含每个节点的出度个数拷贝的集合,最后对该集合进行随机匹配,匹配的节点间相连.

1999 年,美国圣玛丽亚大学(University of Notre Dame)物理系自组织系统研究组的 Barabási 与 Albert[13]在对无尺度网络进行研究时提出了幂律产生的两个系统根源:增长(growth)和优先附着(preferential attachment)(例如,新加入节点优先依附于出度高的节点).前一个特点表明无尺度网络可以不断地扩张,后一个特点则意味着两个节点连接能力的差异可以随着网络的扩张而增大,即“富者愈富”.BA,ESF 和 GLP 都根据这一理论通过模拟网络发展过程来构造拓扑图.在实现优先附着时,3 个模型采用线性优先,即一个新加入节点与图中已存在节点 v 相连接的概率 L 与该节点出度 d_v 的大小线性相关.

$$L(v)=(d_v-c)/\sum_j(d_j-c)\ ,\ c\le 1 \tag{2}$$

© 中国科学院软件研究所 http://www.jos.org.cn

其中,j 属于图中已存在节点集合;c 为常数,其值越小,优先附着性越弱.

BA(Barabási-Albert)model[13]是 Barabási 和 Albert 在提出幂律产生的两个原因的同时提出的一种生成算法.在该算法中 c=0,先从少量的 m_0 个节点开始,每隔一段时间增加一个出度为 $m(\le m_0)$的新节点.

ESF(extended scale free model)[14]是 Barabási 和 Albert 于 2000 年提出的 BA 模型的扩展版本.模型中加入了新的 Internet 发展特征:在网络发展过程中,已有连接会发生变化,即重新连接(rewiring).算法中 c=1,设 4 个常数:$P(0<P<1)$,$Q(0<Q<1-P)$,正整数 m 和初始的 m_0 个节点.过程为 3 步,循环反复:第 1 步,以概率 P 增加 m 条边;第 2 步,以概率 Q 对 m 条边重新连接;第 3 步,以概率 $1-P-Q$ 增加一个有 m 条边的新节点.

GLP(generalized linear preference)[15].2002 年,Tian Bu 与 Towsley 针对 PLRG,BA 和 ESF 生成的拓扑图与真实 Internet 拓扑图在“small-world”理论中的特征路径长度和聚类系数上存在差异,提出了一种新的拓扑生成算法.在该算法中,c 为参数,以由 m_0-1 条边相连的 m_0 个节点开始,每一步进行下面两个操作之一:(1) 以概率 p 增加 $m(\le m_0)$条边;(2) 以概率 $1-p$ 添加一个出度为 m 的新节点.

图取样(map sampling)[16]法是 Magoni 和 Pansiot 在 2002 年提出来的.图取样法生成的拓扑图不是“凭空”制造的,而是以真实的 Internet 拓扑图加上指定的节点数量和平均出度作为输入参数,再从 Internet 拓扑图中随机选取一幅子图,按照算法从子图中随机抽取部分节点,最后输出一幅符合 Internet 拓扑特征的路由器级拓扑图.

4 拓扑生成器

拓扑生成器是拓扑生成算法的软件实现,设计目标可归纳为 3 点:(1) 代表性,即生成的人工拓扑图要能够准确反映实际 Internet 拓扑的各个方面;(2) 包含性,即将许多生成模型合并在一起,提供一个通用的拓扑生成工具;(3) 兼容性,即为网络模拟应用程序(如 ns-2)以及可视化工具(如 CAIDA 的 Otter)提供接口.为了实现这些目标,一个拓扑生成器主要包括 3 个功能模块:(1) 格式转换.由于目前没有统一的拓扑数据文件格式标准,所以拓扑生成器要具有在各种拓扑测量数据,拓扑生成器输入输出数据以及网络模拟软件输入数据之间进行文件格式转换的功能.(2) 拓扑生成.这是拓扑生成器的核心.除了根据已有的拓扑生成算法实现幂律之外,还要依据其他度量来产生拓扑图.(3) 拓扑分析.根据拓扑生成模块中所采用的度量来对拓扑数据进行统计分析,给出度量的值.目前,专门的拓扑生成器有以下 3 种:

BRITE(Boston University representative Internet topology generator)[17]是波士顿大学于 2001 年开发的一个通用拓扑产生器.其特点在于“通用”:首先,它实现了 Waxman 模型和 BA 算法,并可以用自顶向下和自底向上两种方法来产生 GT-ITM 层次模型;其次,它为多个网络模拟程序提供接口,包括 ns-2,SSF,Omnet++,JavaSim,并支持可视化工具 Otter.BRITE 产生路由器级和 AS 级拓扑图,并可为连接分配带宽和延迟,具有图形用户接口.

Inet[18]是密歇根大学于 1999~2002 年间开发的一个 AS 级拓扑产生器,最新发布的是 3.0 版本.其特点是利用 route-views.oregon-ix.net 上 1997 年 11 月~2002 年 2 月间的大量 BGP 数据来确定拓扑度量的值,可信性很高.Inet 采用 PLGR 算法与优先附着实现幂律,重视连通性(最小节点覆盖),并针对最大团尺寸和聚类系数做了优化.Inet 使用简便,只需在命令行输入拓扑图中节点总数即可.

nem(network manipulator)[19]是法国 ULPS 大学(University of Louis Pasteur Strasbourg)于 2002 年开发的.其特点是具有通用性,并对拓扑图在最短路径长度和树结构方面的度量进行优化.nem 实现了除早期 3 种算法与 GLP 之外的全部 5 种拓扑生成算法.nem 将拓扑特征分为 5 类:树、距离、连通性、节点对最短路径数以及类别(class,包括叶上节点数、环上节点数等),可对这 5 类特征进行分析.nem 为网络模拟软件 ns-2,GloMoSim 和 OPNet 提供接口,以批处理文本文件作为用户接口.

不包括幂率模型,上面一共介绍了 12 种能够产生拓扑图的拓扑模型、拓扑生成算法或拓扑生成器,那么它们之中哪一个最接近 Internet 拓扑结构呢?可以这样说,在它们之中每一个被提出来时,都与当时已经存在的模型或算法进行了分析比较,以证明自己更优秀.也就是说,对这些模型或算法的评价可以参照它们的“生日”来进行.已发表的最新比较工作是 2002 年 Magoni 和 Pansiot[27]对 6 个具有代表性的网络拓扑生成器的 15 项指标进行的测评.测评指标涵盖了幂律、路径长度和多播树相关度量.满分为 40 分,测评成绩从高到底为 nem(38 分),ESF 与 Inet2.1(33 分),Model A 与 PLOD(31 分),BRITE(21 分).可见,目前的拓扑模型和产生器可以产生很接

© 中国科学院软件研究所 http://www.jos.org.cn

近 Internet 的拓扑图了.

5 难点问题与研究路线

Internet 拓扑研究目前面临的主要困难来自于两个方面:一方面是 Internet 自身复杂的特性;另一方面是缺乏进行系统研究的方法论.

Internet 自身性质带来的问题主要集中在拓扑测量上,具体表现在:拓扑结构动态变化影响测量准确性;其庞大的规模影响了测量数据的完整性;异构性与管理的非集中性增加了测量的技术难度[1].因此客观地说,目前的网络测量技术无法获得完整而准确的 Internet 拓扑图,现有的 Internet 拓扑模型都存在一定的局限性.由此引发的问题是:Internet 拓扑中真的存在幂律吗?这个疑问的产生来自于两点:(1) Qian Chen 等人[20]发现 Faloutsos 等人[13]在发现幂律时所用的 AS 拓扑数据丢失了 20%~50%的物理连接,与真实的网络拓扑之间存在很大差异,也就是说,幂律并不能严格地刻画 Internet 拓扑;(2) Lakhina 等人[21]也对幂律的正确性提出了挑战,指出由少数探测源点和庞大的目标节点构成的探测体系由于最短路径原理的作用而存在采样偏见(sampling biases).文章指出,采用这种存在采样偏见的技术对 Waxman 模型产生的拓扑图进行测量,实验结果同样会满足幂律.另外,Internet 的动态性也缩短了拓扑模型中静态元素的寿命,使得拓扑建模研究不能仅仅描述当前 Internet 的表象,还要发现其背后的驱动机制.由此引发了另一个问题:Internet 拓扑中的幂律是如何形成的?对于这个问题,Qian Chen 等人[20]提出,历史拓扑数据不能支持 BA 模型中提出的基于连通性动力学对 Internet 发展所作的假设,建议从 HOT(highly optimized tolerance)理论中寻求答案.

随着研究的不断深入,人们发现 Internet 不只是硬件加上软件,而是在一定意义上进行“新陈代谢”的自组织系统.Internet 拓扑与万维网(通过 URL 连接)、细胞网络(通过化学反应连接)、人际关系网、科学文献引用网络等具有共同的内在特征——自相似性,并且认识到对 Internet 拓扑缺乏了解并不只是一个计算机科学问题,而是源于缺少一个对复杂网络进行特征化的科学框架[10].因此,成功建立 Internet 拓扑模型要依赖于科学界在上述领域的共同努力,拓扑建模的研究成果也会促进这一科学框架的形成.

综上所述,今后 Internet 拓扑建模的研究路线主要包括以下 3 个方面:(1) 通过分析路由协议与协议栈来开发新的拓扑测量技术,提高数据完整性以及路由器多接口合并比例,以求从新数据中发现新特征;(2) 将数据挖掘与图论两者结合起来,从现有数据中发现新特征;(3) 借鉴其他复杂网络的研究成果,特别是复杂系统中具有普遍性的原理,以发现 Internet 拓扑形成的内在机制.

References:

[1] Floyd S, Paxson V. Difficulties in simulating the Internet. IEEE/ACM Trans. on Networking, 2001,9(4):392~403.

[2] Zheng H. Internet worm research [Ph.D. Thesis]. Tianjin: Nankai University, 2003 (in Chinese with English abstract).

[3] Chalmers RC, Almeroth KC. On the topology of multicast trees. IEEE/ACM Trans. on Networking, 2003,11(1):153~165.

[4] NMS Home Page. http://www.darpa.mil/ipto/programs/nms/index.htm

[5] Zhang HL, Fang BX, Hu MZ, Jiang Y, Zhan CY, Zhang SF. A survey on Internet measurement and analysis. Journal of Software, 2003,14(1):110~116 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/14/110.htm

[6] Waxman BM. Routing of multipoint connections. IEEE Journal on Selected Areas in Communications, 1988,6(9):1617~1622.

[7] Doar MB. A better model for generating test networks. In: Proc. of the GLOBECOM’96. London: IEEE, 1996. 86~93.

[8] Zegura EW, Calvert KL, Donahoo MJ. A quantitative comparison of graph-based models for Internet topology. IEEE/ACM Trans. on Networking, 1997,5(6):770~783.

[9] Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the Internet topology. ACM SIGCOMM Computer Communication Review, 1999,29(4):251~262.

[10] Albert-László Barabási. The physics of the Web. 2001. http://www.physicsWeb.org/article/world/14/7/09

[11] Palmer CR, Steffan JG. Generating network topologies that obey power laws. In: Proc. of the GLOBECOM 2000, Vol 1. San Francisco: IEEE, 2000. 434 ~438.

© 中国科学院软件研究所 http://www.jos.org.cn

[12] Aiello W, Chung F, Lu LY. A random graph model for massive graphs. In: Proc. of the ACM STOC 2000. Portland: ACM Press, 2000. 171~180.

[13] Barabási AL, Albert R. Emergence of scaling in random networks. Science, 1999,286(5439):509~512.

[14] Albert R, Barabási AL. Topology of evolving networks: local events and universality. Physical Review Letters, 2000,85(24):5234.

[15] Tian Bu, Towsley D. On distinguishing between Internet power law topology generators. In: Proc. of the IEEE INFOCOM 2002, Vol 2. New York: IEEE, 2002. 638~647.

[16] Magoni D, Pansiot JJ. Internet topology modeler based on map sampling. In: Proc. of the ISCC 2002. Taormina: IEEE, 2002. 1021~1027.

[17] Medina A, Lakhina A, Matta I, Byers J. BRITE: An approach to universal topology generation. In: Proc. of the MASCOTS 2001. Washington: IEEE Computer Society, 2001. 346~353.

[18] Jared Winick, Sugih Jamin. Inet-3.0: Internet topology generator. Technical Report, CSE-TR-456-02, Ann Arbor: University of Michigan, 2002.

[19] Magoni D. nem: A software for network topology analysis and modeling. In: Proc. of the MASCOTS 2002. IEEE Computer Society, 2002. 364 ~371.

[20] Qian Chen, Hyunseok Chang, Govindan R, Jamin S. The origin of power laws in Internet topologies revisited. In: Proc. of the IEEE INFOCOM 2002, Vol 2. New York: IEEE, 2002. 608~617.

[21] Lakhina A, Byers JW, Crovella M, Xie P. Sampling biases in IP topology measurements. In: Proc. of the IEEE INFOCOM 2003, Vol 1. San Francisco: IEEE, 2003. 332~341.

[22] Tauro SL, Palmer C, Siganos G, Faloutsos M. A simple conceptual model for the Internet topology. In: Proc. of the GLOBECOM 2001, Vol 3. San Antonio: IEEE, 2001. 1667~1671.

[23] Magoni D, Pansiot JJ. Analysis of the autonomous system network topology. ACM SIGCOMM Computer Communication Review, 2001,31(3):26~37.

[24] Watts D, Strogatz S. Collective dynamics of ‘small-world’ networks. Nature, 1998,393(6684):440~442.

[25] Tangmunarunkit H, Govindan R, Jamin S, Shenker S, Willinger W. Network topology generators: Degree-Based vs. structural. ACM SIGCOMM Computer Communication Review, 2002,31(4):147~159.

[26] Gkantsidis C, Mihail M, Zegura E. Spectral analysis of Internet topologies. In: Proc. of the IEEE INFOCOM 2003, Vol 1. San Francisco: IEEE, 2003. 364 ~374.

[27] Magoni D, Pansiot JJ. Evaluation of Internet topology generators by power law and distance indicators. In: Proc. of the IEEE ICON 2002. Singapore: IEEE, 2002. 401~406.

附中文参考文献:

[2] 郑辉. Internet 蠕虫研究[博士学位论文].天津:南开大学,2003.

[5] 张宏莉,方滨兴,胡铭曾,姜誉,詹春艳,张树峰.Internet 测量与分析综述.软件学报,2003,14(1):110~116. http://www.jos.org.cn/1000-9825/14/110.htm

© 中国科学院软件研究所 http://www.jos.org.cn