

AS 级 Internet 拓扑度和连通性演化分析

付大愚 赵海葛新

(东北大学信息科学与工程学院 沈阳 110004)

摘要 Internet 拓扑,尤其是 AS 级拓扑,是目前研究的热点。研究 Internet 拓扑的演化趋势,可以更好地了解网络的内在连接机制。该项研究基于 CAIDA 项目授权的海量数据,数据采集时间为 2004 年 1 月至 2008 年 6 月。首先介绍了必要的基本概念,然后给出了节点平均度、最大节点度值、高度值节点平均度、富人俱乐部连通性和聚集系数的演化分析。研究发现,度值较高的节点,其影响力随时间变化逐渐下降,高度值节点之间的连接也逐渐减少,但网络整体上仍表现出明显的富人俱乐部性质和聚集性。

关键词 AS 级,Internet 拓扑,度,连通性,演化分析

中图分类号 TP393 **文献标识码** A

AS-level Internet Topology Degree and Connectivity Analysis

FU Da-yu ZHAO Hai GE Xin

(College of Information Science and Engineering, Northeastern University, Shenyang 110004, China)

Abstract The Internet topology, especially the AS-level topology, is the hotspot of the research. We can comprehend well the inner connective mechanism of the network by researching the evolvement trend of the Internet topology. In this thesis, the research task is based on the massive data authorized by CAIDA(The Cooperative Association for Internet Data Analysis) Skitter project and the data's time span is from January 2004 to June 2008. This paper introduced the essential basic conceptions first, then carried out the evolvement analysis of the node average degree, the maximum node degree, the average degree of the top-degree node, the connectivity of rich-club and the clustering coefficient. It is discovered that the affect of the top-degree nodes and the connectivity between the top-degree nodes is descended by the time movement. However the network puts up still obvious rich-club character and clustering coefficient totally.

Keywords AS-level, Internet topology, Degree, Connectivity, Evolvement analysis

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,文献[1]的研究表明,Internet 的节点数大约每两年翻一番。如今 Internet 的“面貌”已与其原型 ARPANET 大相径庭,成为一个由计算机构成的“复杂自组织生态系统”。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 拓扑研究就是探求在这个看似混乱的网络之中蕴含着哪些还不为人们所知的规律。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。

1 AS 级拓扑分析和定义

对 Internet 的研究自 Internet 诞生开始^[2-4],一直是层出不穷,经久不衰。在早期,人们更多关注的是 Internet 的体系结构、网络协议^[5]、计算机互联以及 Internet 所提供的服务等方面的研究。特别是近几十年来人们在复杂性科学和复杂网络等领域取得的研究成果,使国内外的研究者认识到 Internet 也是复杂网络之一,由此人们开始了从复杂性和复杂网络的

角度对 Internet 进行研究。

近年来人们在该领域取得了长足的进展,发现了许多隐藏在网络内部的特征规律。但目前的相关研究工作存在数据统计的空间量级较小,数据分析时间跨度较短,度量方法较为简单等问题,所以需要更为全面的进一步研究。

Internet 的路由选择结构是一种层次式的选择结构,由若干路由器汇集成一个 AS(Autonomous System,自治系统),不同 AS 之间再通过边界路由器彼此互连^[6]。所以,Internet 拓扑的研究工作也主要集中在 AS 级和路由级两个层面开展。由于与路由级拓扑相比,AS 级拓扑位于网络的更“上”一层,其特征与变化对 Internet 的影响更为巨大,相关研究对网络未来的发展意义更为重大。同时,AS 级拓扑数据集规模小,以现有计算能力来说,可以更加有效执行深层次、高时间复杂度的计算分析,探究更为深层次的客观特性。根据 AS 拓扑分析的结果,可以总结更为系统的分析策略以及可靠的分析手段,以利于进一步发现路由级拓扑中的隐含规律,为超大规模网络统计分析提供可用手段。

Internet 是动态变化的,其拓扑也是随时间而改变的,所

到稿日期:2008-11-24 返修日期:2009-02-16 本文受国家自然科学基金资助项目(69873007),国家级火炬计划项目(2002EB010154)资助。
付大愚(1972-),男,博士研究生,主要研究方向为计算机网络、复杂网络等,E-mail:fdymail@yeah.net;赵海(1959-),男,教授,博士生导师,主要研究方向为复杂网络、数据融合等;葛新(1980-),男,博士研究生,主要研究方向为复杂网络。

以研究 Internet 拓扑的演化趋势,可以更好地了解网络的内在连接机制。下面介绍本文的几个重要概念^[7,8]。

定义 1(图, Graph) 一个图 G 是一个三元组,这个三元组包括一个顶点集 $V(G)$ 、一个边集 $E(G)$ 和一个关系,该关系使得每一条边和两个顶点(不一定是不同的点)相关联。这两个顶点称为这条边的端点。

从定义中可以看到,从任意顶点 x 到 y 不能连接两条或两条以上的边。本文所讨论的图,均符合上述要求,即均为不含多重边的图。

定义 2(网络, Network) $N=(V, E, c, X, Y)$ 为一个网络,如果

(1) $G=(V, E)$ 是一个有向图;

(2) c 是 E 上正整数,称为容量函数,对于每条边 $e, c(e)$ 称为边 e 的容量;

(3) X 与 Y 是 V 的两个非空不相交子集,分别称为 G 的发点集与收点集, $I=(X|\overline{X}\cup Y)$ 称为 E 的中间点集, X 的顶点称为发点(源), Y 的顶点称为收点(汇), I 的顶点称为中间点。

在图论中,网络 $G=(V, E)$ 是指由点集 $V(G)$ 和边集 $E(G)$ 组成的图,且 $E(G)$ 中的每条边 e_i 有 $V(G)$ 中的一对点 (u, v) 与之对应。记顶点数为 $N=|V|$, 边数为 $L=|E|$ 。如果任意 (u, v) 与 (v, u) 对应同一条边,则称为无向网络,否则称为有向网络;如果任意 $|e_i|=1$, 则称为无权网络,否则称为加权网络。

定义 3(度, Degree) 设 N 是一个网络, $V(N)$ 是所有顶点的集合, $E(N)$ 是所有边的集合。顶点 v 的度 k_v 是指与此顶点 v 连接的边的数量, $v \in V$ 。即

$$k_v = \sum_{l \in E} \delta_l^v$$

其中,当边 l 包含顶点 v 时, δ_l^v 取值为 1, 否则为零,即

$$\delta_l^v = \begin{cases} 1, & \text{if } l \text{ include } v \\ 0, & \text{if not} \end{cases}$$

定义 4(节点平均度, node average degree) 网络 N 中所有节点 i 的度 k_i 的平均值,即:

$$\bar{k} = \frac{1}{N} \sum k_i$$

定义 5(簇系数, Clustering coefficient) 图的簇系数是衡量图的成团特性的参数。对于无向图 $G(V, E)$, 记某顶点 x 的邻点集合为 $A(x)$, 显然 $d(x) = |A(x)|$, 顶点 x 的簇系数定义为它所有相邻节点之间连边的数目占可能的最大连边数目的比例,即:

$$C_G(x) = \frac{2e(G[A(x)])}{d(x)(d(x)-1)}$$

图 G 的簇系数则被定义为所有顶点簇系数的平均值:

$$C(G) = \frac{1}{v(G)} \sum_{x \in V} C_G(x)$$

定义 6(富人俱乐部, Rich-club) Internet 中有少量的节点具有大量的边,这些节点被称为“富节点(rich nodes)”。这些富节点倾向于彼此之间相互连接,构成“富人俱乐部”。

定义 7(富人俱乐部连通性, Rich-club connectivity) 可以用富人俱乐部连能性 $\Phi(r/N)$ 来刻画富人俱乐部现象,它表示的是网络中前 r 个度最大的节点之间实际存在的边数 L 与这 r 个节点之间总的可能存在的边数 $r(r-1)/2$ 的比值,即:

$$\Phi(r/N) = \frac{L}{r(r-1)/2} = \frac{2L}{r(r-1)}$$

如果 $\Phi(r/N)=1$, 那么前 r 个最富的节点组成的富人俱乐部为一个完全连通的子图。

2 度特征量演化分析

CAIDA(The Cooperative Association for Internet Data Analysis)是一个对全球范围 Internet 结构及数据进行研究的国际合作机构,CAIDA Skitter 是由全世界范围的主要研究机构与高等学府参与并涉及多个领域和交叉学科的大型科研项目,研究的主要内容包括 Internet 网络的产生、发展及演化趋势,以及 Internet 网络行为、动力特征和 Internet 宏观拓扑结构的变化规律。Skitter 通过跟踪从一个源地址到多个目的地址的前向 IP 地址的方式来获取 Internet 的拓扑结构,通过收集 ICMP 协议的 TTL 值(生存时间)数据来绘制 Internet 节点间关系。东北大学嵌入式技术实验室经该组织授权,成立 CAIDA 中国第一节点(Neu node),成为该组织在中国的首家合作伙伴,共享研究成果,并保持长期经验技术交流。本文选取 2004 年 1 月至 2008 年 6 月 CAIDA Skitter 项目提供的 AS 级拓扑的最新测量数据进行分析,对网络特征量进行大时间跨度演化分析,观察网络生长过程中各项指标的变化,并分析其未来趋势。还选取多种统计量进行分析,尽可能避免由于观察角度单一而导致的特征遗漏。

本节对所选用的数据集做初步演化分析,主要包括以下常规特征量。

(1) 节点平均度

图 1 给出 AS 级 Internet 节点平均度值随时间的变化规律。横轴为实际测量数据的时间跨度,纵轴为节点平均度值。图中平均度值在 6.4 至 5.5 之间振荡变化,但振幅较小。整体趋势是节点平均度值随网络增长而下降。

(2) 最大节点度值

图 2 给出 AS 级 Internet 拓扑中节点度值的最大值随时间的变化规律。横轴为实际测量数据的时间跨度,纵轴为最大节点度值,其值抖动范围较大,但整体趋势是下降的,说明网络中最“热”节点(度值最大的节点)的影响力随时间变化逐渐下降。

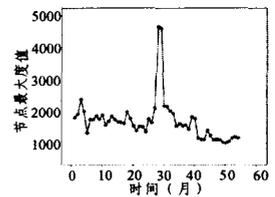
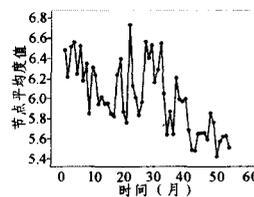


图 1 节点平均度随时间的统计变化图 图 2 最大节点度值随时间的统计变化图

3 连通性演化分析

网络的连通情况对于其性能的影响是至关重要的,而网络中的节点通过之间的连接互通,所以连接的变化可以导致网络连通情况的变化。分析网络中节点度值的变化,可以有效地发现其中连接的演变趋势。图 3 给出的是 AS 网络中度值最大的 0.5% 与 1% 的节点集合内,其节点度值的平均值随

(下转第 140 页)

- [2] Lee D, Yannakakis M. Principles and methods of testing finite state machines-a survey [C] // Proceedings of The IEEE 84. 1996;1090-1123
- [3] 李华,叶新铭. 基于 Petri 网的测试路径选择标准[J]. 内蒙古大学学报:自然科学版,2007,38(1):88-93
- [4] Li hua, Ye Xinming. Generating Executable Test Sequences Based on Petri-net for Combined Control and Data Flow of Communication Protocol [C] // ICCT'98 International Conference. 1998
- [5] Parrish A S, Zweben S H. On the Relationships Among the Alluses, All-DU-Paths, and All-Edges Testing Criteria [J]. IEEE Transaction on Software Engineering, 1995, 21(12)
- [6] Chen W H. Executable test sequence for the protocol control and data portions [C] // Proc. of IEEE Int'l Conference on Communications. New Orleans, U. S. A. , 2000
- [7] Miller R E, Paul S. Generating Conformance Test Sequences for Combined Control and Data Flow of Communication Protocols [C] // Proceeding of Protocol Specifications, Testing and Verification (PSTV'92). Florida, USA, June 1992
- [8] 张玉军,李忠诚. 移动 IPv6 测试中的层次化协议描述和测试生成方法[J]. 电子学报, 2004, 32(12A): 30-34
- [9] 施新刚,尹霞. 移动 IPv6 协议一致性测试研究与实践[J]. 计算机工程, 2007, 33(11), 116-118
- [10] 赵达观,谢高岗,张大方. Mobile IPv6 Home Agent 协议一致性测试集的设计与实现
- [11] Hang Y S, Ural H U. Dependence Testing; Extending Data Flow Testing with Control Dependence [M]. Berlin, Heidelberg: Springer, Volume 3502, 2005
- [12] 刘静. 基于 TTCN-3 的移动 IPv6 协议一致性测试研究[D]. 呼和浩特:内蒙古大学, 2008

(上接第 105 页)

时间演化的情况。

图 3 中显示度值较高的两个节点集平均度随时间缓慢下降,这与图 1 及图 2 的情形类似,即高度值的节点随网络整体度值的下降,自身度值也在下降。图中二者变化曲线的形状十分接近,与图 1 相比也很类似,这说明网络中平均度值的下降主要是由于高度值节点的度值下降所致。

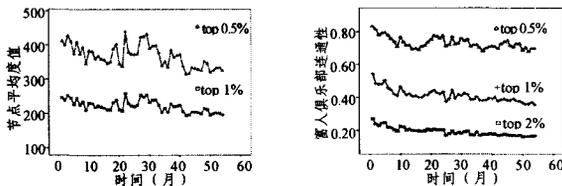


图 3 高度值节点平均度随时间的统计变化图 图 4 富人俱乐部连通性随时间的统计变化图

网络中存在度值较高的“富”节点,富人俱乐部性质体现了其相互之间具有很大的连接倾向性,富人俱乐部连通性则对该性质做了量化度量;其值越大,“富”节点间的联系越紧密。考察 AS 网络中度值最大的 0.5%, 1% 以及 2% “富”节点集,图 4 给出其各自富人俱乐部连通性的演化趋势。三个节点集的连通性均随时间变化有所下降,且下降趋势基本一致,说明该现象是受到同一因素影响所致。

富人俱乐部性质主要是由高度值节点间的连接倾向形成的,其连通性下降说明此类节点间的连接逐渐减少,结合图 3 所示结论可知,高度值节点集合整体平均度值下降,同时伴随着集合内连接的减少。依文献[9]所述,网络中低度值节点的邻居大部分是高度值节点,而高度值节点表现出紧密连接的富人俱乐部性质,这使得低度值节点的局部聚集系数很高,从而提高了整个网络的聚集系数。

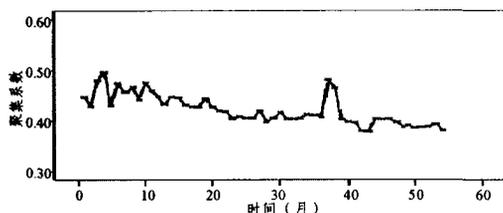


图 5 聚集系数随时间的统计变化图

随着高度值“富”节点间连接的减少,富人俱乐部连通性降低,网络的聚集系数也应有所下降。图 5 给出聚集系数随时间的变化情况,可见其下降趋势十分明显,验证了之前的分析。

连通性分析表明,AS 网络中高度值节点之间的连接在逐渐减少,并由此导致了网络中无论是“富”节点之间还是普通邻居节点之间的连通性均有所下降,但网络整体仍表现出明显的富人俱乐部性质与聚集性。

结束语 通过对 AS 级 Internet 拓扑特征量的演化分析发现,度值大节点对于其它节点的影响力在下降,也可以说节点间连接能力的差距呈减小趋势。与此同时,网络中度值大节点之间以及普通邻居节点之间的连通性也有所下降。

参考文献

- [1] Willinger W, Doyle J. Robustness and the Internet: Design and evolution [EB/OL]. <http://netlab.caltech.edu/Internet/>, 2002
- [2] Li J, Sung M, Xu J, et al. Large-scale IP traceback in high-speed Internet; practical techniques and theoretical foundation [C] // Proceedings of the IEEE Symposium on Security and Privacy. California, USA, 2004
- [3] Subramanian L, Padmanabhan V N, Katz R H. Geographic properties of Internet routing [C] // Proceedings of the USENIX Annual Technical Conference. 2002
- [4] Akella A, Seshan S, Balakrishnan H. The impact of false sharing on shared congestion management [C] // Proceedings of the 11th IEEE International Conference on Network Protocols, 2003
- [5] Jose M, Barcelo, Juan I, et al. Study of Internet autonomous system interconnectivity from BGP routing tables [J]. Computer Networks; The International Journal of Computer and Telecommunications Networking, 2004, 45(3): 333-344
- [6] 王大东,王洪军,王瑞军,等. 一种基于 AS 的 Internet 拓扑模型 [J]. 计算机工程, 2005, 31(4): 23-25
- [7] West D B. 图论导引 [M]. 北京:机械工业出版社, 2006: 1-47, 339-348
- [8] Munkress J R. 拓扑学 [M]. 北京:机械工业出版社, 2006: 2-52
- [9] 张国强,张国清. Internet 网络的关联性研究 [J]. 软件学报, 2006, 17(3): 490-497