

# 基于多点测量的网络节点度分布研究<sup>\*</sup>

李超 赵海 张昕 葛新  
(东北大学信息科学与工程学院 沈阳 110004)

**摘要** 首先简要描述了单点测量网络存在的采样偏见问题,认为通过增加探测源节点,可以最大限度地降低网络连接边遗漏的采样偏差。在此基础上,以 CAIDA 分布于世界范围内的多个探测源点得到的采样拓扑数据,分别考察了网络的节点度三种不同的规律分布。frequency-degree 分布表明实际的网络节点度分布服从幂律,但幂指数随着采样偏见影响的降低呈减小的趋势,degree-rank 服从幂律分布说明了网络节点的度分布存在着较高的不均匀性,补累积函数表明 Weibull 分布对节点度分布的拟合效果更好。

**关键词** Internet 测量,拓扑特征,幂律,谱密度

## Research on the Distribution of Network Nodes Based on Multiple Measurement Points

LI Chao ZHAO Hai ZHANG Xin GE Xin

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

**Abstract** Sampling bias caused by few points was first present, and by increasing the measurement points can reduce the omission of network edge. Based on the giant data samples authorized by CAIDA, the distribution of network nodes was analyzed respectively. The distribution of frequency-degree presents that the nodes of network follows power-law, and the distribution of network nodes exists heterogeneity confirmed by the degree-rank distribution. Cumulative function indicates that Weibull distribution may be more suitable to describe the distribution of network nodes.

**Keywords** Internet measurement, Topology characteristic, Power-law, Spectral density

### 1 引言

随着网络测量数据的获取,Internet 拓扑研究经历了从早期的经验假设逐渐转向客观分析的过程<sup>[1]</sup>。但是,网络自身的复杂性以及采样数据手段的不完善,将对网络拓扑结构的描述以及网络拓扑建模会产生一定的影响<sup>[2]</sup>。

一般来说,根据测量点的个数,网络测量可以分为单点测量与多点测量<sup>[3]</sup>。对于由少数探测源节点和庞大的目标节点构成的探测体系将存在采样偏见<sup>[4]</sup>,单个探测节点对网络拓扑结构测量的结果往往与网络实际形态有较大差距。假设目标网络存在一个 traceroute 探测服务器, a, b, c, d 叶子节点,其余节点为路由器。traceroute 服务器对目标网络进行主动测量,网络测量结果如图 1(a)所示。

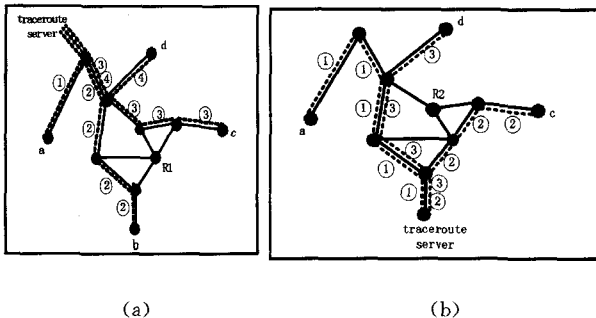


图 1 单点测量目标网络

从图 1(a)看到,路由器  $R_1$  以及与  $R_1$  相连接的四条链路无法探测到,其主要原因在于通往目标叶子节点的路由器及路由连接存在冗余,造成在探测目标叶子节点时,某些冗余路由及连接的缺失。同样,当叶子节点 b 为 traceroute 服务器,尽管目标叶子节点全部被探测到,但探测路径与网络实际拓扑结构存在较大差异。

如果从图中两个探测点开始同时测量目标网络,那么目标网络的探测结果如图 2 所示。

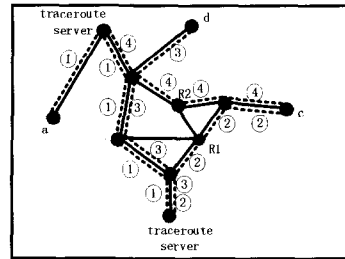


图 2 两点测量目标网络

从图 2 看到,两个 traceroute 服务器将图 1 中探测到的路由数及连接进行合并,图 1 中遗漏的路由器  $R_1, R_2$  均被测量得到。因此通过增加探测源节点,可以在一定程度上减小由于网络中探测源节点较少引起的路由节点遗漏和网络连接边的采样偏差。

<sup>\*</sup> 国家自然科学基金资助项目(69873007),国家级火炬计划项目(2002EB010154)。李超 博士生;赵海 教授,博士生导师。

1. CAIDA (The Cooperative Association for Internet Data Analysis), 是一个对全球范围内 Internet 结构及数据进行研究的国际合作机构,其主要目的是在全球范围合作基础上最大限度地对 Internet 数据进行获取、分析、研究和共享。东北大学嵌入式技术实验室已经成为 CAIDA 在中国授权建立的第一监测节点。

本文通过 CAIDA<sup>1</sup> skitter<sup>[5]</sup>项目授权的真实海量样本数据,从 CAIDA 分布于世界范围内 20 多个监测节点得到的采样结果作为分析样本,其数据量巨大,可靠性高,在最大限度地降低采样偏差的基础上,本文对 Internet 的网络拓扑结构作初步分析。

## 2 网络样本数据

为使测量结果真实地反映网络特征,本文以随机选取初始点,逐跳采样相邻路由由节点之间连接边的方式,对测量数据处理,得到以初始节点为核心的局部 Internet 拓扑结构。

分别以 1, 2, 5, 21 个 CAIDA 监测点对 Internet 进行采样,所测量到的 IP 总数、总边数以及相应的平均边数如表 1 所示。

表 1 CAIDA 21 个探测点对 Internet 测量结果

探测点数	探测总 IP 数	总边数	平均边数
1	361166	711516	3.5095
2	412842	1092480	3.9283
5	600435	2099270	4.4663
21	1307423	5434716	4.8677

从表 1 可以看出,随着探测点数量增加,可探测到 Internet 的 IP 个数、总边数均呈现明显增长。对 CAIDA 探测点来说,由于使用的探测方法相同且探测目标集(即目标 IP 地址集)相同,因此 IP 地址总数的增加主要是由更多的监测点探测到了更多的路由器,并且表 1 表明平均度的变化范围不大,能够大致反映网络拓扑结构特征,可以对其分析。

## 3 网络节点度分布

### 3.1 frequency-degree 分布特征

通过分析多点测量得到的 Internet 拓扑结构度分布能够反映网络拓扑结构特征<sup>[6]</sup>,观察不同监测点采样得到的网络拓能够降低存在的采样误差,并且可以得出监测节点数目对网络拓扑特征描述产生的影响。

本文分别以 1 个监测点、2 个监测点、5 个监测点以及 21 个监测点所观测得到的 Internet 拓扑子图以及合并图作为样本数据集,对网络拓扑的节点度(degree)及其频数(frequency)分布做拟合分析,结果如图 3 所示。

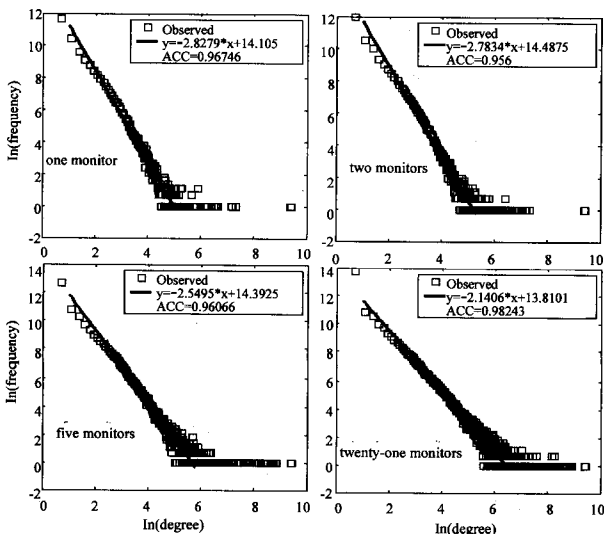


图 3 Internet 测量拓扑图的 frequency-degree 幂律分布及拟合图(对数-对数坐标)

观察图 3 可以看到,4 个子图均表现了良好的幂律特性,表 2 反映了幂指数数值变化的规律,其相关系数绝对值 ACC 在 0.95 以上,说明拟合直线与观测数据拟合程度较好。

表 2 frequency-degree 幂律分布幂指数变化情况

探测点数	ACC	幂指数( R )
1	0.9675	2.8279
2	0.9560	2.7834
5	0.9601	2.5495
21	0.9824	2.1406

从图 3 可以看到,在不同监测点下,节点频数与节点度之间均符合较为明显的幂律分布,并且表 2 表明,当监测节点数逐渐增加后,幂指数呈逐渐减小的趋势。因此,通过增加监测节点的个数,可以在较大程度上降低由于单点测量 Internet 导致的幂律偏向问题。

### 3.2 degree-rank 分布特征

网络节点频数在一定程度上服从幂律分布表明较少的节点具有较高度,大量节点其度数较低,说明 Internet 网络拓扑结构具有较高的非均一性。针对节点度与节点等级(rank)之间的关系进行分析,可以对网络节点度的层次结构进行描述。

将 4 个拓扑特征实例的拓扑子图对节点度大小按照降序(rank)排列,得到的节点度与该节点等级的 degree-rank 分布及拟合情况如图 4 所示。

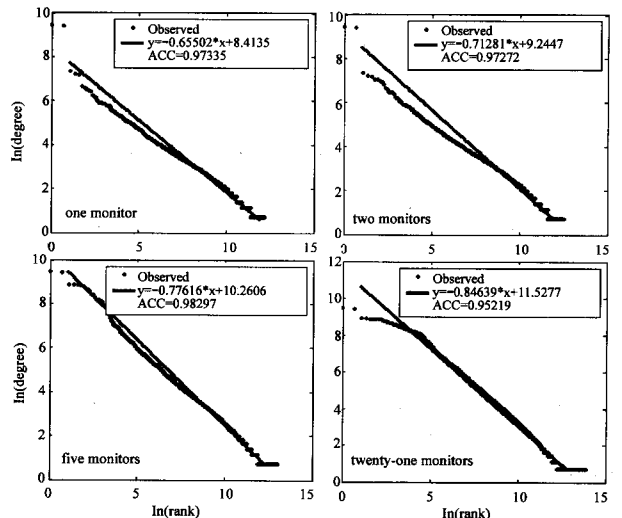


图 4 Internet 测量拓扑图的 frequency-degree 幂律分布及拟合图(对数-对数坐标)

degree-rank 服从幂律分布说明了 Internet 节点的度分布存在着较高的不均匀性,较大度值节点个数与较小度值节点个数之间相差较大。从表 3 看到相关系数绝对值 ACC 在 0.97 以上,说明拟合直线与观测数据的拟合程度较好。

表 3 degree-rank 幂律分布幂指数及度比值变化情况

探测点数	ACC	幂指数( R )	最小度个数/次 小度个数
1	0.9734	0.6550	3.3921
2	0.9727	0.7128	4.2578
5	0.9830	0.7762	6.7064
21	0.9941	0.8464	17.4633

且锁的相似考虑只会增加锁请求被授予的概率。

### 参考文献

[1] Lam K Y, Kuo T W, Tsang W H. Transaction Shipping Approach for Mobile Distributed Real-time Databases. Database and Expert Systems Applications, 1999; 932-941  
 [2] Barbara S. A multi-version data model for executing real-time transactions in a mobile environment // Proceedings of the 2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access. 2001; 90-97  
 [3] Lam K Y, Kuo T W, Tsang W H, et al. Concurrency Control in Mobile Distributed Real-time Databases. Information Systems, 2000, 25(4): 261-286  
 [4] Lam K Y, Kuo T W, Tsang W H, et al. A Similarity-based Protocol for Concurrency Control in Mobile Distribution Real-time Database Systems // IPPS/SPDP Workshops. 1999; 329-338

[5] Lee V C Y, Lam K-Y, Kao B. Priority Scheduling of Transaction in Distributed Real-time Databases. Journal of Real-time Systems, 1998, 15(1): 31-61  
 [6] Lai C. Optimistic Similarity-based Real-time Concurrency Control // Proceedings of IEEE Real-Time Technology and Applications Symposium, Denver, IEEE Computer Society Press, 1998; 189-198  
 [7] Kayan E, Ulusoy O. An Evaluation of Real-time Transaction Management Issues in Mobile Database Systems. The Computer Journal, 1999, 42(6): 501-510  
 [8] Kayan E, Ulusoy O. Real-time Transaction Management in Mobile Computing Systems. Database Systems for Advanced Applications, 1999  
 [9] Spyridon, Braoudakis, et al. Concurrency Control Protocols For Real-time databases. Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, 1995

(上接第 99 页)

随着监测节点数增加,在幂指数绝对值增大的同时,最小度个数与次小度个数的比值明显增大。因此在保证拟合准确的前提下,拟合直线的斜率变小,即向绝对值变大方向偏转,从而导致了等级幂指数的变化。

### 3.3 CCDF(d)-degree 分布特征

文献[7]指出频率  $f_d$  的补累积分布函数(complementary cumulative distribution function, 简称 CCDF)与度  $d$  也具有幂律关系,称该幂律为 CCDF(d)-degree 幂律。分别以 1 个监测点, 2 个监测点, 5 个监测点及 21 个 CAIDA 监测点所观测到的数据进行拟合, 结果如图 5。

表 4 CCDF(d)-degree 幂律分布的补累积分布函数拟合结果

探测点数	函数形式	标准残差平方和(SSSE)
1	Power law	12455.6927
	Power law(2)	219431.0825
	Weibull(2-parameter)	11594.8785
2	Power law	24215.0629
	Power law(2)	303397.4291
	Weibull(2-parameter)	20133.3965
5	Power law	114594.8493
	Power law(2)	503785.6687
	Weibull(2-parameter)	59191.7273
21	Power law	485010.9747
	Power law(2)	1160172.4009
	Weibull(2-parameter)	221809.1604

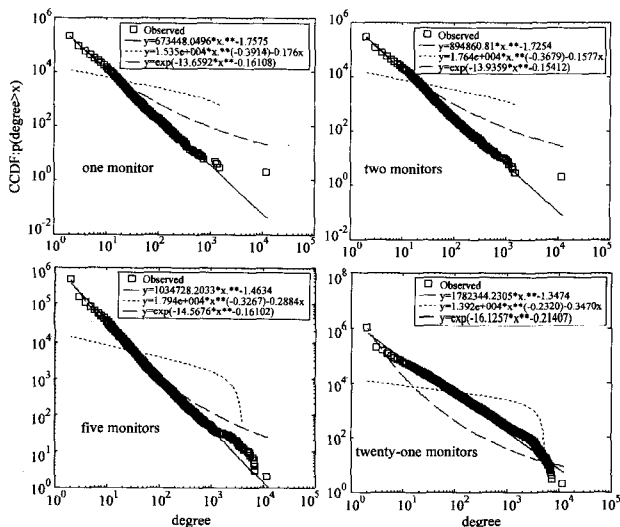


图 5 Internet 测量拓扑图的 CCDF(d)-degree 幂律分布及拟合图(对数-对数坐标)

从图 5 直观地看到,监测节点观测数据的补累积分布函数与节点度在对数坐标下呈直线变化关系,表明网络拓扑节点度的补累积分布服从幂律分布。为量化 3 种分布函数对实际观测数据的拟合优度,以标准残差平方和(SSSE)对 3 种分布的拟合优度进行评价。

从表 4 看到 Power law(2)在 4 个子图中的拟合优度均大于 Power law,说明与 Power law(2)的补累积分布函数形式  $F'(x) = -\frac{C}{\alpha+1}x^{\alpha+1} + Dx$  相比,Power law 的补累积分布  $F'(x) = -\frac{C}{\alpha+1}x^{\alpha+1}$  更适合表达 Internet 拓扑特征。并且 Power law 补累积分布的标准残差平方和大于 Weibull(2-parameter)的标准残差平方和,因此,就节点度的分布情况,Weibull 分布比幂律分布拟合效果更好。

**结束语** 本文通过对 CAIDA 组织授权的海量样本数据进行采样,对其节点度分布进行分析,发现随着监测节点数的增加,采样偏见影响地降低,幂指数呈逐渐减小的趋势。degree-rank 服从幂律分布说明了网络节点的度分布存在着较高的不均匀性。通过比较 3 种拟合函数的标准残差平方和发现,Weibull(2-parameter) 分布的拟合优度均优于 Power law 分布。因此,就节点度的分布情况,Weibull 分布可能比幂律分布更适于体现 Internet 的网络度分布。

### 参考文献

[1] 张宇,张宏利,方滨兴. Internet 拓扑建模综述[J]. 软件学报, 2004, 15(8)  
 [2] Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the Internet topology[J]. ACM SIGCOMM Computer Communication Review, 1999, 29(4): 251-262  
 [3] 张宏莉,方滨兴. Internet 测量与分析综述. 软件学报, 2003, 14(1)  
 [4] Lakhina A, Byers JW, Crovella M, et al. Sampling biases in IP topology measurements[A] // Proc. of the IEEE INFOCOM 2003, Vol 1. San Francisco, IEEE, 2003; 332-341  
 [5] CAIDA skitter Project. <http://www.caida.org>  
 [6] 姜誉,方滨兴,胡铭曾. 大型 ISP 网络拓扑多点测量及其特征分析实例[J]. 软件学报, 2005, 16(5)  
 [7] Siganos G, Faloutsos M, Faloutsos P, et al. Power laws and the AS-level Internet topology[J]. IEEE/ACM Trans on Networking, 2003, 11(4): 514-524