

Internet 平均路径长度的定量研究

赵金晶¹, 朱培栋², 牛薇¹, 吴侃², 曹鸿强³

(1. 北京系统工程研究所, 北京 100101; 2. 国防科学技术大学计算机学院, 长沙 410073; 3. 北京图形研究所, 北京 100029)

摘要: 探讨了自治系统级 Internet 平均路径长度的变化趋势问题。研究了理论模型及近 5 年的真实网络数据, 发现两者中得到的 Internet 平均路径长度的变化趋势完全相反。分析了产生这种现象的原因, 通过建模得到了计算真实情况下 Internet 平均路径长度的公式, 在此基础上对 2006 年~2012 年 Internet 的平均路径长度变化趋势进行了预测。

关键词: 自治系统级 Internet; 平均路径长度; 幂律; 小世界

Quantitative Research on Average Path Length of the Internet

ZHAO Jin-jing¹, ZHU Pei-dong², NIU Wei¹, WU Kan², CAO Hong-qiang³

(1. Beijing Institute of System Engineering, Beijing 100101; 2. School of Computer, National University of Defense Technology, Changsha 410073; 3. Beijing Graphic Research Institute, Beijing 100029)

【Abstract】 This paper focuses on the trend of the average path length of AS-level Internet. The average path trends of the Internet drawn by the academic models and its real data of the recent 5 years are different. It anatomizes the reason and gets the better formulation of the average path length in real environment by modeling, and forecasts the trend of the average path length from 2006 to 2012.

【Key words】 AS-level Internet; average path length; power-law; small world

对BGP数据和实时测量数据进行分析发现, Internet拓扑存在着幂律(power-law)特征: $P(k) \sim k^{-\gamma}$, $2 < \gamma < 2.5$ 。相对负指数分布, Internet自治系统互联度数的概率分布曲线下落得比较缓慢, 表现出拖尾(heavy-tailed)特点^[1]。域间路由系统同时具有小世界(small world)的特性, 平均AS-path比较短, 路由信息在AS之间传递的平均跳步不大。在此基础上产生了各种刻画Internet的结构模型^[2-6]。在网络中, 两点间的距离被定义为连接两点的最短径所包含的边数, 求所有节点对的距离平均值就得到了网络的平均路径长度。它是衡量网络转发能力的重要参数。具有较短的平均路径长度可以降低传输延迟、减少代价。因此, 它在很多相关的研究领域(如路由、查找、信息传输^[7]等)相当重要。

1 平均路径长度模型

在 power-law 网络和 small world 网络中, 平均路径长度都远小于随机网络。

1.1 small world 模型

Watts 和 Strogatz 于 1998 年提出一个小世界模型(WS)模型, WS 模型的生成规则是在规则网络中随机加入少量的长程边, 这样的得到网络拓扑具有较大聚集系数和较短最短路径。

平均路径长度 l 可以定义为网络中任意两点的平均最短路径长度:

$$l(p) = \frac{1}{2} \frac{\sum_{i,j} d(p)_{ij}}{n(n-1)} \quad (1)$$

其中, d_{ij} 为从节点 i 到节点 j 的几何距离; p 为演化概率。当 $p=0$ 时, 为随机网络, 平均路径长度 $l \sim N/2K$, K 为常数; 当 $p=1$ 时, $l \sim \ln(N)/\ln(K)$, 即随着节点 N 的增加、网络规模的扩大, 平均路径长度 l 随 N 呈对数增长。

1.2 power-law 模型

给定一个节点, 从它开始找到最近的、次最近的、……、第 m 个邻居, 假设网络中所有节点都能够在 l 步内到达, Z_m 为第 m 步所到达的平均节点数, 那么,

$$1 + \sum_{m=1}^l Z_m = N \quad (2)$$

在 power-law 网络中, 节点度分布满足幂律性:

$$P(k) = Ck^{-\gamma}, \quad \text{for } k \geq 1 \quad (3)$$

根据式(2)和式(3), 平均路径长度 l 可以被定义为

$$l = \frac{\ln(n) + \ln[\xi(\gamma)/\xi(\gamma-1)]}{\ln[\xi(\gamma-2)/\xi(\gamma-1)]} + 1 \quad (4)$$

式(4)表明, l 随网络规模 n 呈对数关系。

因此在 small world 和 power-law 网络模型中, 平均路径长度的增长等于或者小于节点数的对数。从理论分析来看, l 呈慢增长趋势。

2 Internet 平均路径长度的数据统计

为了观察 Internet 平均路径长度的变化趋势, 从 Routeviews 中采集了 2001 年~2006 年的 BGP 路由表, 从中提取出 AS-path 属性。

图 1 是 AS1221 在 2001 年~2006 年从 20 000 个采集点上采集的 AS 平均路径长度的值, 其平均值为 3.461 1。可以看出, 平均路径长度的值呈平稳下降的趋势, 这与第 1 节中模

基金项目: 国家自然科学基金资助项目(60673169); 国家“863”计划基金资助项目(2006AA01Z213)

作者简介: 赵金晶(1981-), 女, 博士, 主研方向: 域间路由技术; 朱培栋, 副教授、博士; 牛薇, 工程师、硕士; 吴侃, 本科生; 曹鸿强, 高级工程师

收稿日期: 2007-03-30 **E-mail:** pdzhu@nudt.edu.cn

型分析得出的理论值完全相反。

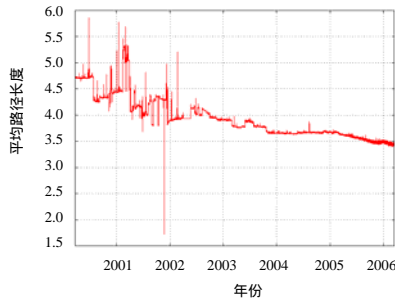


图1 AS-path的平均长度分布

作为 Internet 中一个独立的基本单元, AS 通过连接组成了 AS 级网络。ISP 级的拓扑建模和产生需要考虑经济因素和技术因素。ISP 的经济约束包括它能够支持的物理连接, 即接纳流量和输出流量代价的平衡, 因此, ISP 拓扑模型需要考虑这种经济影响因素。技术因素主要考虑物理和硬件的条件, 如接口和线卡的数目等, 还要考虑地理位置的影响。当一个新的 AS 需要连接进网络时, 它需要考虑以下的因素来选择连接的邻居: (1)地理位置; (2)商业关系、经济因素成为调节的杠杆; (3)技术约束、接口板、带宽要求等; (4)一定程度的冗余度来保证服务的质量, 可能需要连接到多个 AS 上, 即多宿主。这些因素都使得 AS 级的拓扑更加复杂和不可预知。因此, 真实情况和单纯的模型分析结果之间产生了分歧。换句话说, Internet 是一个符合 power-law 特性和 small world 特性的复杂网络。

3 Internet 平均路径长度预测

Internet 的成长模式使得平均路径长度变小。这种变化看起来是随机且不可预测的, 但是通过之前对数据的分析, 发现它的变化也满足一定的规律。

Internet 遵从 power-law 特性, 具有以下参数:

定义 1 hop-plot 指数 H : h 个跳步内可以到达的节点对数与跳数 h 的 H 次方成正比:

$$P(h) \propto h^H, \quad h < \text{Diameter}(G) \quad (5)$$

定义 2 秩指数 R : 节点 v 出度 d_v 与节点的秩 r 的 R 次方成正比:

$$d_v \propto r^R \quad (6)$$

定义 3 邻居平均规模 $NN(h)$: 在 h 个跳步内可得到达的邻居数。

根据 power-law 特性有

$$NN(h) = \frac{P(h)}{N} - 1 = \frac{N+2E}{N} h^H - 1 \quad (7)$$

其中, $NN(h)$ 即式(2)中参数 Z_m 的累加和。则平均路径长度 l 可以由式(2)和式(7)推导得到:

$$1 + NN(l) = N; \quad \frac{N+2E}{N} l^H = N$$

$$l = \left(\frac{N^2}{N+2E} \right)^{1/H} \quad (8)$$

根据图的幂率特性, 边数 E 与节点数 N 存在如下关系:

$$E = \frac{N}{2(R+1)} \left(1 - \frac{1}{N^{R+1}} \right) \quad (9)$$

其中, R 为图 G 的秩指数。式(9)的具体证明过程可以参见文献[8-9]。则由式(8)和式(9)可得

$$\delta ef = \left(\frac{N^2}{N+2E} \right)^{1/H} = \left(\frac{N^2}{N + \frac{N}{R+1} \left(1 - \frac{1}{N^{R+1}} \right)} \right)^{1/H} = \left(\frac{(R+1)N^{R+2}}{(R+2)N^{R+1} - 1} \right)^{1/H} \quad (10)$$

这样, 可以根据 R, N, H 的值对有效直径进行预测。其中, AS 的数量 N 在 2001 年之前是以指数级速率增长的, 而从 2001 年~2007 年的近 6 年时间内以近线性增长, 见图 2。 R 与 H 从 1997 年~2005 年的变化曲线如图 3 所示。

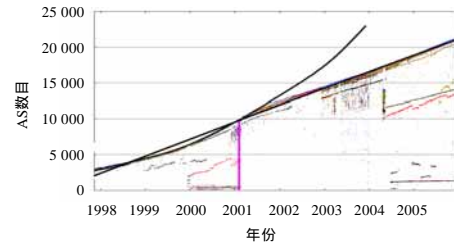
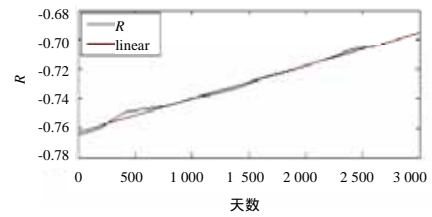
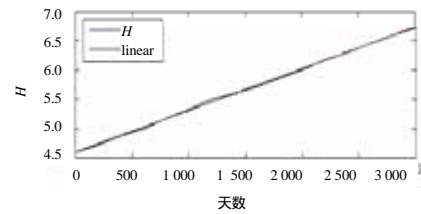


图2 AS数目增长曲线



(1)



(2)

图3 R与H的变化曲线与线性拟合的误差

为了得到 N, R, H 随时间的变化关系, 对这 3 条曲线进行线性拟合, 得到以下方程 (t 以天为单位):

$$N = 6 \times t + 82 \quad (11)$$

$$R = 2.26 \times 10^{-5} t - 0.76 \quad (12)$$

$$H = 7.12 \times 10^{-4} t + 4.6 \quad (13)$$

那么通过将式(11)~式(13)代入式(10)就可以得到 δef 与 t 的关系函数如下:

$$\delta ef = \left(\frac{(2.26 \times 10^{-5} t + 0.24) \times (6t + 82)^{(2.26 \times 10^{-5} t + 1.24)}}{(2.26 \times 10^{-5} t + 1.24) \times (6t + 82)} \right)^{\frac{1}{(7.12 \times 10^{-4} t + 4.6) - 1}} \quad (14)$$

图 4 是 2006 年~2012 年有效直径的变化曲线。可以看出, 有效直径会逐年缩小, 速率接近于 0.000 25, 在 2012 年, Internet 的平均路径长度将达到 3.0。

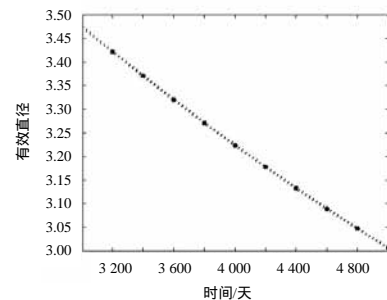


图4 有效直径随N与H的变化曲线

(下转第 162 页)