

Skitter 与 Ark 探测架构下 AS 级 Internet 拓扑分析

张君¹ 赵海¹ 康敏²

(东北大学信息科学与工程学院 沈阳 110004)¹ (沈阳市气象局 沈阳 110168)²

摘要 选取 CAIDA 授权的 AS 级 2003 年 1 月—2007 年 12 月的 Skitter 数据及 2008 年 1 月—12 月的 Ark 数据进行层层深入的对比分析,以说明 Internet 拓扑探测架构的改变对拓扑探测结果的影响。首先统计了 Skitter 及 Ark 架构下 AS 级 Internet 拓扑的多种宏观特征,分析了 Skitter 与 Ark 在各特征值上的异同,进而分析了网络的幂率特征与网络连通性以及拓扑核数的演化,指出幂率性质以及高聚类性质在 Internet 拓扑中是真实存在的,不随探测方式的改变而消失。

关键词 复杂网络, Skitter, Ark, 幂率, 连通性, AS 级拓扑

中图分类号 TP301.5, TP393 **文献标识码** A

Analysis of Internet AS-level Topology under Skitter and Ark Measurement Infrastructure

ZHANG Jun¹ ZHAO Hai¹ KANG Min²

(College of Information Science & Engineering, Northeastern University, Shenyang 110004, China)¹

(Shenyang Meteorological Administration, Shenyang 110168, China)²

Abstract This paper selected Internet AS-level Skitter dataset from Jan, 2003 to Dec, 2007 and Ark dataset from Jan, 2008 to Dec, 2008 authorized by CAIDA (The Cooperative Association for Internet Data Analysis) to analyze in-depth to explain the effect on the measurement results of the Internet topology by the changing of the measurement infrastructure. Research began with the detailed statistics of various characteristics of Internet AS-level topology. The similarities and differences in several characteristics of Internet AS-level topology under Skitter and Ark were analyzed. Then the power-law distribution and the evolution of connectivity and topology core of the network were analyzed. The results show that the power-law property and high clustering feature of Internet have not disappeared with the transition of infrastructure.

Keywords Complex network, Skitter, Ark, Power law, Connectivity, AS-level topology

互联网作为一个典型的复杂网络实例,对其宏观拓扑结构的特征分析及建模研究具有重要意义,受到学术界的广泛关注。近年来人们在该领域的研究取得了长足的进展,尤其是在自治系统层面 (autonomous system, AS-level)^[1-5]。2008 年初,CAIDA (The Cooperative Association for Internet Data Analysis) 的宏观拓扑测量项目组用 Ark 架构代替了原始的 Skitter 架构,新的探测手段在拓扑数据上带来了哪些变化,这些变化所反映出的互联网特征与原有结论是否一致,这些都是需要继续研究的问题。

1 Skitter 与 Ark

CAIDA 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 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。

CAIDA 的 Skitter 项目于 1998 年开始运行,主要用于探测全球的 Internet 拓扑,并进行相关的性能测量。Skitter 架

构采用集中式测量,各探测源点定时从位于美国圣地亚哥的 CAIDA 中心服务器下载目标列表,使用 traceroute 方法探测网络路径,探测结束后将结果返回给中心服务器。在 Skitter 架构下各探测源点是各自为政,相互独立的。

2008 年 2 月 8 日,CAIDA 的宏观拓扑测量项目组正式用新型的拓扑探测架构 Archipelago (简称 Ark) 代替运行了 10 年之久,收集了近 4 千兆数据的 Skitter。Ark 作为对 Skitter 的改进,保留了 Skitter 的测量源点、相关测量工具等,同时架设了一台中间 Web 服务器来分配目的列表,一台文件存储服务来收集测量点探测到的路径,并提供一个公共 Web 服务器方便协作者进行数据下载。Ark 与 Skitter 最大的不同是使用元组空间来代替原来的通信模块,用 scamper 代替了原来的 Skitter 测量工具。

新型的探测方法通过元组空间进行信息传递及命令控制,使得各测量源点间可进行协作性工作,形成了网络测量的分布式平台。Ark 使用组探测方式调整了基于 traceroute 的大规模拓扑测量方法,使其可以在短时间内覆盖更大的地址空间,更敏锐地捕捉到互联网的变化。在从 Skitter 迁移到

到稿日期:2009-12-23 返修日期:2010-03-09 本文受国家自然科学基金项目(60973022)资助。

张君(1967-),女,博士,讲师,主要研究方向为复杂网络,E-mail:zhangjun@ise.neu.edu.cn;赵海(1959-),男,教授,博士生导师,主要研究方向为复杂网络、普适计算、嵌入式系统及数据融合等;康敏(1963-),女,工程师。

Ark 之后,其探测所覆盖的范围更加完整,仅 AS 级数据集的规模就翻了一番。新的探测手段得到的 Internet 拓扑更接近真实的 Internet,那么人们原来在 Skitter 架构上所进行的研究以及得到的结论是否仍然成立呢?为此我们展开了下面的研究。

2 Skitter 与 Ark 架构下 Internet 演化分析

为了保证分析所用数据在时间维度上的冗余性,本文选取 Internet 2003 年 1 月到 2007 年 12 月的 Skitter 数据以及 2008 年 1 月到 12 月的 Ark 数据进行分析。

2.1 常规特征量分析

2.1.1 网络规模演化

图 1 给出了 AS 级 Internet 拓扑节点数与连接数随时间的演化规律。

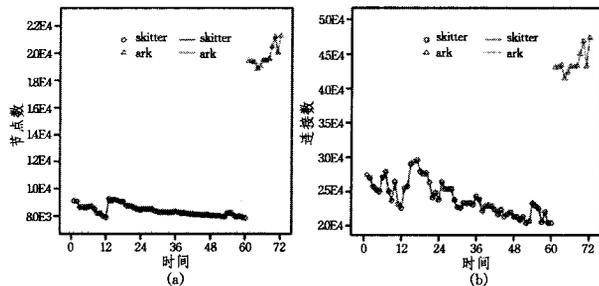


图 1 节点数与连接数的演化

节点数在一定程度上体现了网络规模的大小,图 1(a)为 AS 级 Internet 拓扑节点数目随时间的演化规律。横轴为实测数据的时间跨度,共 72 个月。其中包括 60 个月的 Skitter 数据及 12 个月的 Ark 数据;纵轴为节点总数,即 Internet 中 AS 的数目。从图中可以看出 Skitter 的数据在 2004 年初存在一个快速上升阶段,而后逐渐下降。随着时间的推移,节点数目稳定在 8000 左右。而 Ark 数据呈现震荡上升,节点数目在 18000 以上。

连接数是衡量网络规模的另一个量,它表示节点间连接的总边数。图 1(b)为 AS 级 Internet 拓扑连接数随时间的演化情况。横轴为实测数据的时间跨度,纵轴为连接数。从图中可见,Skitter 的连接数在 2 万至 3 万之间震荡变化,整体呈现下降趋势;Ark 的连接数在 4 万以上。

从网络规模上来看,Ark 的节点数及连接数是 Skitter 的两倍左右。这说明与 Skitter 相比,Ark 发现了更多的隐藏节点与连接,其可探测到的范围更广。

2.1.2 叶子节点变化

叶子节点亦即度值为 1 的节点,图 2 为 AS 级 Internet 拓扑叶子节点数目及比例随时间的演化情况。

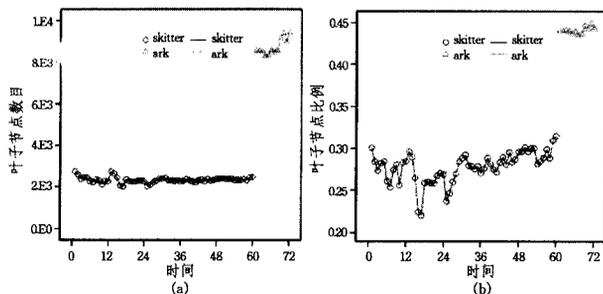


图 2 叶子节点数目及比例的演化

图 2(a)为 AS 级 Internet 拓扑叶子节点数目随时间的演化情况。横轴为实测数据的时间跨度,纵轴为叶子节点的数目。Skitter 的叶子节点数目稳定在 2000 至 4000 之间,Ark 的叶子节点数目大于 8000。这是由于 Ark 的探测范围较大,发现边缘节点的能力优于 Skitter。

图 2(b)为 AS 级 Internet 拓扑叶子节点比例随时间的演化情况。Skitter 的叶子节点比例主要在 20% 至 30% 之间;而 Ark 的叶子节点比例维持在 45% 左右。这进一步说明了 Ark 探测边缘节点的能力优于 Skitter。

2.1.3 整体度值演化

平均度是网络拓扑的基本连接特性,网络的平均度越大表示其平均连接程度越高。同时,这个网络可能具备更好的鲁棒性。图 3 给出了 AS 级 Internet 拓扑的平均度随时间的演化情况,其横轴为实际测量数据的时间跨度,纵轴为节点平均度值。Skitter 的平均度值在 5.0 至 6.5 之间变化,其震荡幅度不大,整体呈下降趋势;Ark 的平均度值变化极小,稳定在 4.5 左右,这是由于 Ark 探测边缘节点的能力优于 Skitter,从而导致网络拓扑的平均度略低于 Skitter 探测的结果。

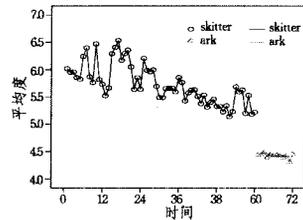


图 3 平均度的演化

2.2 幂率分析

节点度的幂率分布是 Internet 的一个重要规律^[6],在 AS 级拓扑中表现十分明显。它可以体现网络中边的分布情况,即少数节点拥有大量连接,而大多数节点的连接数很低。幂指数可以量化说明这种拓扑的扭曲现象。

定义 1 (degree-rank 幂率) 其分布形式的数学定义如下:

$$d_v \propto r_v^{\beta}$$

式中, d_v 是节点 v 的度, r_v 是将网络中的节点按度降序排列节点 v 的序列号, R 是秩指数常数。

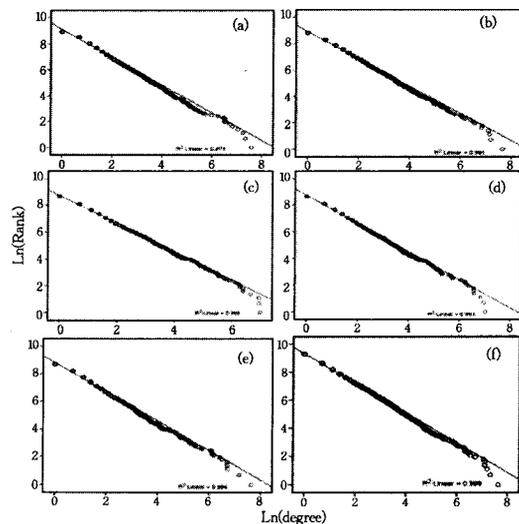


图 4 秩与度值的关系

为考察 AS 级拓扑的幂率特性,选择了 2003 年 6 月到

2008年6月六个时间点的数据,计算其幂率形式与度值的关系如图4所示,其横轴表示节点度值 d ,纵轴表示度值 d 的秩,二者均取自然对数值并拟合了分布曲线斜率。子图(a)至(f)分别为2003年至2008年对应时间点的数据,其中(a)至(e)为Skitter数据集结果,(f)为Ark数据集结果,可见6幅子图均呈现幂率分布。对应的拟合秩幂指数依次为0.913,0.958,0.945,0.938,0.931,0.942,Skitter与Ark的秩幂指数相一致。这说明探测架构从Skitter迁移到Ark后,虽然网络覆盖范围更广,网络规模等特征量发生了较大变化,但Internet的幂率特性并未因此发生改变。这表明Internet的幂率特性不是特定的探测架构产生的现象,所以也不会随着探测架构的迁移而消失。

2.3 连通性分析

2.3.1 富人俱乐部连通性

富人俱乐部性质^[7]体现了网络“富”节点相互之间的连接倾向性程度,其值越大,“富”节点间的联系越紧密。图5为AS级Internet拓扑的富人俱乐部连通性随时间的演化情况。

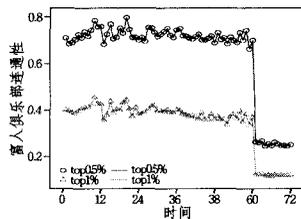


图5 富人俱乐部连通性的演化

从图中可以看出,Skitter的富人俱乐部现象略有下降,由于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。Ark的富人俱乐部连通性变化幅度很小,意味着对应top节点集内部的连接变化不大。由图可见,Ark的富人俱乐部现象不如Skitter明显。

2.3.2 聚类系数

聚类系数^[8]反映了邻居节点之间的相互连接情况,它表达了拓扑本身的鲁棒性:一个节点的聚类系数越高,其邻居节点间连通性越强,从而丰富了该节点的路径多样性。聚类性越强的网络可以使一些路由策略性能更佳。图6为AS级Internet拓扑的聚类系数随时间的演化情况。从图中可以看出,Skitter的聚类系数随时间逐渐下降,而Ark数据的聚类系数变化不明显。Ark的聚类系数虽然低于Skitter,但是仍然保持在一定的高度,这说明Internet拓扑中存在着较高的聚集性。

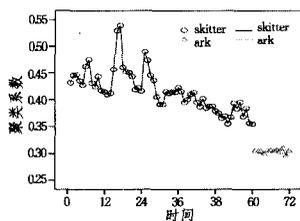


图6 聚类系数的演化

2.4 网络拓扑核数演化

图7给出了AS级Internet拓扑的网络拓扑核数^[9]随时间的演化情况,可以看出无论是Skitter还是Ark数据,其网

络拓扑核数都变化不大,Ark的网络拓扑核数略低。这表明在AS级Internet拓扑中存在着明显的层次性质。

网络拓扑核数的演化分析应更深入地考虑最高核的演化情况,图8为AS级Internet拓扑最高核所含节点数随时间的演化情况。

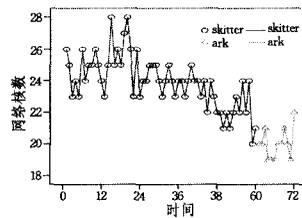


图7 网络拓扑核数的演化

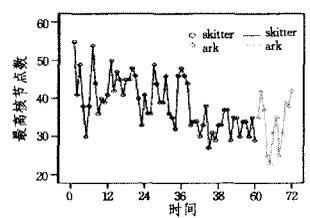


图8 最高核节点数演化

由此可见,Skitter数据早期变化幅度较大,而后逐渐稳定。值得注意的是,虽然Ark的节点总数近似Skitter的两倍,但其最高核节点数目却处于同一数量级。这表明Internet AS级拓扑中处于核心地位的节点数目是稳定的。

结束语 本文针对CAIDA对Internet拓扑探测架构的改变,分析了Skitter及Ark数据规律,发现Ark架构较之Skitter更具发现边缘节点的能力;Ark虽然在网络覆盖规模上优于Skitter,发现了更多的隐藏节点与连接,但仍符合Internet拓扑的幂率性质以及较高的聚类性质,这说明了Internet的幂率性质和高聚集性并不是某种探测架构采样偏见造成的,不随探测方式的转变而消失。

参考文献

- [1] Jose M, Barcelo, Juan I, et al. Study of Internet autonomous system interconnectivity from BGP routing tables[J]. Computer Networks, 2004, 45(3): 333-344
- [2] Mahadevan P, Krioukov D, Fomenkov M, et al. The Internet AS-level topology: three data sources and one definitive metric[J]. SIGCOMM Comput. Commun. Rev., 2006, 36(1): 17-26
- [3] Vilhar A, Novak R. Policy relationship annotations of predefined AS-level topologies[J]. Comput. Netw., 2008, 52(15): 2859-2871
- [4] Krishnamurthy V, Faloutsos M, Chrobak M, et al. Sampling large Internet topologies for simulation purposes[J]. Comput. Netw., 2007, 51(15): 4284-4302
- [5] Oliveira R V, Zhang B, Zhang L. Observing the evolution of Internet as topology[J]. SIGCOMM Comput. Commun. Rev., 2007, 37(4): 313-324
- [6] 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
- [7] Zhou S, Mondragon R J. The rich-club phenomenon in the Internet topology[J]. IEEE Communication Letters, 2004, 8(3): 180-182
- [8] Watts D J, Strogatz S H. Collective dynamics of small-world networks[J]. Nature, 1998, 393: 440-442
- [9] Alvarez-Hamelin J I, Dall'Asta L, Barrat A, et al. K-core decomposition of Internet graphs: hierarchies, self-similarity and measurement biases[J]. Networks and Heterogeneous Media, 2008, 3(2): 371-393