

文章编号:1001-9081(2006)10-2275-03

端到端路径容量测量工具性能分析

黎文伟¹, 张大方², 杨金民², 谢高岗³

(1. 湖南大学 计算机与通信学院, 湖南 长沙 410082; 2. 湖南大学 软件学院, 湖南 长沙 410082;
3. 中国科学院 计算技术研究所, 北京 100080)
(liww@hnu.edu.cn)

摘要:分析已有路径容量测量工具的性能, 定义了评价工具性能的指标集, 在自行搭建的可控网络环境和可重复的多种背景流量条件下, 对几种典型的测量工具性能进行实验评估。实验发现 pathchar、clink 等基于 VPS 技术的工具测量误差较大; pathrate 准确性虽较高, 但测量开销大。大背景流量下现有工具的准确性、测量开销等都不理想。

关键词:网络测量; 带宽测量; 路径容量; 性能分析

中图分类号: TP393.06 **文献标识码:**A

Performance analysis of end-to-end path capacity measurement tools

LI Wen-wei¹, ZHANG Da-fang², YANG Jin-min², XIE Gao-gang³

(1. College of Computer and Communication, Hunan University, Changsha Hunan 410082, China;
2. School of Software, Hunan University, Changsha Hunan 410082, China;
3. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: The performance of current path-capacity measurement tools were analyzed and an index set for tools performance evaluation was set. And several typical path-capacity measurement tools were evaluated in a self-building controllable network under many repeatable cross traffic load conditions. The main findings include, VPS based tools such as pathchar, clink, have relatively larger measurement errors; although pathrate is more accurate, it has bigger measurement costs. When the cross traffic load is high, all of these tools are inaccurate and need great measurement costs.

Key words: network measurement; bandwidth measurement; path capacity; performance analysis

0 引言

测量 IP 网络端到端路径容量是网络监测的重要内容^[1]。路径容量可用于多媒体业务选择适当的服务器和流传播率, 以及覆盖网和应用层组播业务构建性能更优的覆盖网或组播树等。实用的路径容量测量工具, 准确性是基本要求; 为减少测量对网络的影响并获得有意义的测量结果, 还需要工具保持较小探测量和测量时间。由于网络动态性, 也要在不同流量条件下能保证结果一致。

目前已有很多路径容量测量方法和工具, 如基于可变包长技术(Variable Packet Size, VPS)技术的 pathchar、clink^[2]、pchar^[3]和基于包对技术(Packet Pair Time Dispersion, PPTD)的 pathrate^[4]等。但由于缺少在完全可控、可重复网络环境下对这些工具性能的综合实验和评估, 人们难以选择合适工具, 合理评价测量结果。有研究测量并分析了互联网路径容量分布特征^[5], 但由于未考虑工具的因素, 其结果有待验证。

本文在定义了工具性能评价指标集后, 构建了小型 IP 网, 并仿真真实网络流量条件, 对具有代表性的测量工具进行综合实验评估, 比较工具的性能特征, 并分析其原因。

1 典型工具选择

路径容量被定义为: 在没有背景流量条件下, 某端到端路径能够给数据流提供的最大 IP 层吞吐量。路径中传输速率最小的链路(narrow link)决定了路径容量。若 H 是路径跳数; C_i 是链路 i 的传输速率, 路径容量 C 为:

$$C = \min_{i=1 \dots H} C_i \quad (1)$$

路径容量测量方法主要有 VPS 技术和 PPTD 技术两类。VPS 技术基于单包时延模型, 可测量路径每跳链路容量, 主要工具有 pathchar、clink、pchar 等。其基本思想是: 从源端到路径每跳的最小时延不含探测包的排队时间, 与包长度成正比, 与链路容量成反比; 从路径第一跳开始, 测量不同长度探测包的最小时延并进行回归分析即可获得路径每跳链路容量。

PPTD 技术基于包对时延模型, 主要工具有 pathrate、bprobe^[5]、nettimer^[6]等。其基本思想是: 源端向目的端背靠背地发送两长度相等包组成探测包对, 包对在目的端的接收时间间隔与包长度成正比, 与路径容量成反比; 以此来计算路径容量。

由于 pchar 在实验机器上无法编译, bprobe 只工作于 SGI Irix 环境, 文献[6]已经发现 nettimer 测量精度与 pathchar、clink 相当; 故只选择 pathchar、clink、pathrate 进行实验, 这也使得可以对 VPS 和 PPTD 两种测量技术进行评估。

收稿日期:2006-04-03; 修订日期:2006-06-06

基金项目:国家自然科学基金资助项目(60473031;60403031);国家863计划项目(2005AA121560);湖南省自然科学基金(05JJ30116)

作者简介:黎文伟(1975-),男,湖南沅江人,博士研究生,主要研究方向:网络测试和性能评估;张大方(1959-),男,上海人,教授,博士生导师,博士,主要研究方向:可信系统与网络容错计算;杨金民(1967-),男,湖南宁乡人,副教授,博士,主要研究方向:软件容错、可信网络;谢高岗(1974-),男,浙江衢州人,副研究员,博士,主要研究方向:下一代互联网。

2 工具性能评估指标集

为客观地综合评估路径容量测量工具,明确定义了数个评估指标。

测量误差: 设测量路径容量为 C^m , 实际路径容量为 C , 测量误差为:

$$e_m = |(C^m - C)|/C \times 100\% \quad (2)$$

测量误差越小,工具测量越准确。由于工具所依赖的时延模型受背景流量影响,显然不同背景流量负载下,工具的测量误差不同。

测量一致性: 若进行 n 次测量后,得到的测量值为 C^1, C^2, \dots, C^n , 平均值为 \bar{C} , 测量一致性为:

$$S_m = \left[1 - \sum_{i=1}^n |C^i - \bar{C}| / (n \times \bar{C}) \right] \times 100\% \quad (3)$$

大 S_m 值表明测量的可重复性好,但测量不一定准确。只有具有小测量误差和大 S_m 值的测量工具,其测量结果才可信赖。由于背景流量亦影响 S_m 的取值,以相同背景流量下的 S_m 比较工具的测量一致性。

测量探测量: 完成一次路径容量测量所需要发送的探测包的总的字节数。小的探测量意味着工具测量效率高,以少量的测量开销就得到了结果。探测量小,测量对网络运行的影响就小,工具也适合大规模的使用。

测量持续时间: 完成一次路径容量测量所需要的时间。它包括两部分:容量探测时间和数据统计分析时间。测量持续时间短意味着工具测量实时性好。

3 实验设计

3.1 实验网络

评估实验的网络结构如图 1 所示。该网络由 4 个路由器组成。为考查交换机等 2 层存储转发设备对测量是否有影响,将两个 CISCO 路由器通过交换机连接。端主机也通过交换机连接到网络。所有链路速率均为 100Mbps,故实际路径容量为 100Mbps。路径容量测量工具设计时通常假设网络路径中只有一条 narrow link, 图 1 中,每条链路均为 narrow link。这是因为硬件条件的限制,但这也更能够检验工具在实际网络环境中的测量能力。测量工具安装在端主机 Source, 端主机 Sink 作为测量的目的主机。主机 Source、Sink 配置相同,为: 1.7GHz P4 CPU, 256MB 内存, 3COM 100Mbps 网卡, 操作系统为 RedHat Linux 7.2。

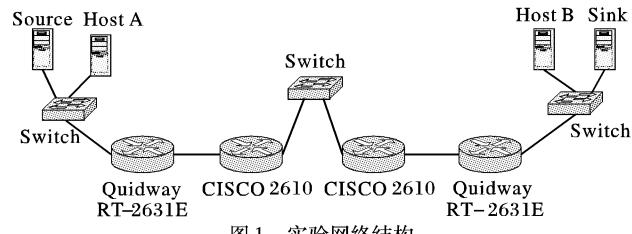


图 1 实验网络结构

路径容量测量工具通常对网络背景流量特征做了某些假设。实际网络环境中,这些假设可能不成立。因此产生尽可能与真实网络流量相似的背景流量是评估实验的关键。由于需要执行多次实验,也要求所产生的背景流量可重复。Harpoon 所产生流量与实际网络流量相似、可重复,并可设定流量速率^[7],故实验中以之作为背景流量产生工具。在图 1 中 Host A 和 Host B 安装了 Harpoon。Host A 做为源端,发送

背景流量;Host B 作为目的端,接收背景流量。背景流量从 0Mbps 递增到 90Mbps,每次增加 10Mbps,相当于链路利用率从 0% 变化到 90%。

3.2 数据收集

为避免端主机及网络中其他因素可能存在的影响,实验时先产生指定速率的背景流量;3min 后启动工具进行容量测量。每次测量结束后,暂停 3min,之后再执行下一次测量。实验期间背景流量始终存在,直至所有工具的测量实验全部完成。每种背景流量条件下,以所选择的工具分别执行了 20 次路径容量测量,取平均值分析。实验时 pathchar、clink 工具记录了每跳链路的容量,以其测量结果中最小值做为所测路径容量。

4 结果分析

4.1 测量误差

表 1 为不同背景流量负载下的测量结果。Pathchar 和 clink 测量结果接近实际路径容量的一半,误差很大。原因是路径中的二层存储转发设备破坏了 VPS 技术的时延模型^[8]。表 2 列出了 pathchar 和 clink 所测量的第二跳链路的容量(链路无二层设备),在背景流量负载不大时,测量结果具有可接收的测量精度。

表 1 不同工具的路径容量测量结果(Mbps)

测量工具	背景流量负载									
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
pathchar	47.8	45.3	43.5	41.7	38.4	32.8	31.7	30.4	31.2	29.8
clink	48.2	46.7	44.1	43.5	41.2	35.4	33.6	32.5	32.1	31.7
pathrate	99.5	98.8	98.4	97.7	94.3	90.6	88.5	87.3	88.3	85.4

表 2 pathchar 和 clink 对路径第二跳链路容量测量结果(Mbps)

测量工具	背景流量负载									
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
pathchar	95.8	92.3	90.2	89.6	83.6	80.7	77.4	78.6	76.1	74.5
clink	97.2	93.4	91.6	89.3	85.4	81.5	76.3	75.7	73.8	72.1

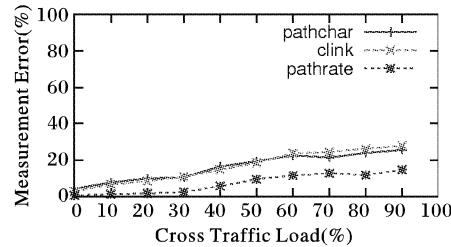


图 2 不同工具的路径容量测量误差

各工具测量误差如图 2 所示。Pathchar 和 clink 以测量第二跳链路容量的误差来比较。背景流量负载达 20% 时,它们的测量误差已接近 10%;而背景流量负载较重时,其测量误差已不可接受(约 25%)。显然基于 VPS 技术对背景流量较敏感,这是因为测量时,VPS 技术是以逐跳链路回归分析来计算链路容量,因而结果包含了较大的累积误差;背景流量负载越大,累积误差也越大。此外,工具记录的用户级时间戳将系统内核处理数据包的时间开销做为了包时延的一部分,这是导致 pathchar 和 clink 测量误差大的另一个原因。背景流量负载较小时(小于 30%),pathrate 测量误差较小,不超过 5%。这是因为 PPTD 技术所依赖的包对时延模型虽然也存在误差,但误差不会累积。此外, pathrate 采用了数据过滤技术来

消除误差较大的测量样本。最后 pathrate 实现了内核级的时间戳记录。故相对而言, 测量误差小。但背景流量负载较重时, 其测量误差也不可忽视。90% 的背景流量负载时, 测量误差约为 15%。由图 2 可知, pathrate 有相对高些的测量准确性, 但所有工具的测量误差随背景流量负载增大而增大。

4.2 测量一致性

各工具的测量一致性如图 3 所示。Pathchar 测量一致性较差, 特别是背景流量较大时。Clink 和 pathrate 采取了自适应样本采集和过滤方法保证测量收敛, 因此其一致性比 pathchar 有较大提高。由图 3 可知, 所有工具的测量一致性随着背景流量的增大而降低。

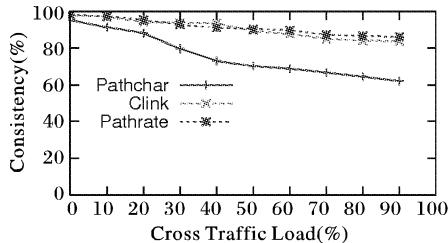


图 3 不同工具的测量一致性

4.3 探测量及测量持续时间

表 3 不同工具的测量探测量

测量工具	背景流量负载									
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
pathchar(KB)	500.5	500.5	500.5	500.5	500.5	500.5	500.5	500.5	500.5	500.5
clink(KB)	266	266	266	266	266	266	266	266	266	266
pathrate(MB)	1.9	2.2	3.7	4.5	5.8	7.3	8.6	10.1	12.6	14.1

表 4 不同工具的测量持续时间(min:sec)

测量工具	背景流量负载									
	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
pathchar	3:43	3:26	3:45	3:52	3:36	3:41	3:29	3:42	3:47	3:38
clink	4:23	4:10	4:02	4:34	4:08	4:12	4:21	4:05	4:14	4:27
pathrate	2:58	3:10	3:48	3:57	4:13	4:37	5:01	5:49	7:04	8:10

各工具的测量探测量和测量持续时间如表 3 和表 4 所示。Pathchar 和 clink 的测量探测量和测量持续时间由测量时的参数决定。我们选择在 50% 背景流量下测量最准确的参数完成其他背景流量负载条件下的测量实验, 故表 3 和表 4 中, 在不同背景流量下其测量探测量和测量持续时间基本

(上接第 2274 页)

法, 找出给定费用下带宽、延迟与延迟抖动满足概率最大的组播路由树。仿真结果表明, 该算法是可行和有效的, 具有较好的性能。

参考文献:

- [1] WANG Z, CROWCROFT J. Quality of service for supporting multimedia application [J]. IEEE Journal on Selected Areas in Communication, 1996, 14(7): 1228–1234.
- [2] SALAMA HF, REEVES DS. Evaluation of multicast routing algorithms for real-time communication on high-speed network [J]. IEEE Journal on Selected Areas in Communications, 1997, 15(3): 332–345.
- [3] CHARIKAR M, NAOR J, SCHIEBER B. Resource optimization in QoS multicast routing of real-time multimedia [J]. IEEE/ACM Transaction on Networking, 2004, 12(2): 340–348.
- [4] ZAPPALA D. Alternate path routing for multicast [J]. IEEE/ACM Transaction on Networking, 2004, 12(1): 30–43.
- [5] 郑彦兴, 田菁, 窦文华. 基于遗传算法的有矢量约束的多播路

无变化。Pathrate 的测量探测量很大, 且随背景流量增大而增长, 其测量持续时间也有类似特征, 这是因为 pathrate 需要发送很多探测包对和包链来过滤测量误差。总体而言, 这几种工具都有较大的探测量开销和较长的测量时间。

5 结语

本文在对可控、可重复的负载条件下对多种路径容量测量工具性能进行了实验评估, 发现: pathchar 较不准确, 测量一致性也差; clink 测量一致性虽好, 但不准确; pathrate 有较小的测量误差和较好的测量一致性, 但其测量探测量和测量时间较多。在背景流量增大时, 所有工具都出现较大的测量误差, 同时测量一致性变差。如何增强路径容量测量工具对大背景流量的适应性是需要解决的问题。

参考文献:

- [1] 汤艳霞, 谢高岗, 雷震甲. 包交换网络带宽的测量技术 [J]. 计算机工程, 2003, 29(17): 46–48.
- [2] DOWNEY AB. Using pathchar to estimate internet link characteristics [A]. Proceedings of ACM SIGCOMM'99 [C]. New York: ACM Press, 1999. 241–250.
- [3] BRUCE AM. Pchar: a tool for measuring internet path characteristic [EB/OL]. <http://www.kitchenlab.org/www/bmam/Software/pchar/>, May 2006.
- [4] DOVROLIS C, RAMANATHAN P, MOORE D. Packet-dispersion techniques and a capacity-estimation methodology [J]. IEEE/ACM Transaction on Network, 2004, 12(6): 963–977.
- [5] CARTER R, CROVELLA M. Measuring bottleneck link speed in packet switched networks [J]. Performance Evaluation, 1996, 27(8): 297–318.
- [6] LAI K, BAKER M. Measuring link bandwidths using a deterministic model of packet delay [A]. Proceedings of ACM SIGCOMM'00 [C]. Stockholm, Sweden: ACM Press, 2000. 283–294.
- [7] SOMMERS J, KIM H, BARFORD P. Harpoon: a flow-level traffic generator for router and network tests [A]. Proceedings of SIGMETRICS 2004 [C]. New York: ACM Press, 2004. 392–392.
- [8] PRASAD RS, DOVROLIS C, MAH BA. The Effect of layer-2 store and forward devices on per-hop capacity estimation [A]. Proceedings of IEEE INFOCOM 2003 [C]. San Francisco, CA: IEEE Press, 2003. 2090–2100.
- [9] JEON Y, KIM J. Application of simulated annealing and tabu search for loss minimization in distribution systems [J]. Electrical Power and Energy Systems, 2004, 26(1): 9–18.
- [10] (日)玄光男, 程润伟. 遗传算法与工程设计 [M]. 汪定伟, 唐加福, 黄敏, 译. 北京: 科学出版社, 2000: 46–59.
- [11] EAGELS PK, NOCOL VM. Recent approaches to global optimization problems through particle swarm optimization [J]. Natural Computing, 2002, 12(1): 235–306.
- [12] Tanenbaum AS. 计算机网络 [M]. 第 4 版. 潘爱民, 译. 北京: 清华大学出版社, 2004: 186–189.
- [13] SRINIVAS N, KALYANMOY D. Multiobjective optimization using nondominated sorting in genetic algorithm [J]. Journal of Evolutionary Computation, 1995, 2(3): 221–248.