带宽测量实验研究及其算法改进

谢高岗[1,3],汤艳霞[2],张大方[1],李忠诚[3]

(1. 湖南大学计算机与通信学院,湖南长沙 410082;2. 西安电子科技大学研究生院,
陕西西安 710071;3. 中国科学院计算技术研究所,北京 100080)

摘 要: 带宽是网络规划、管理和性能优化的重要指标.主动带宽测量技术可以跨越多个自治系统实现端到端的带宽测量,因此被广泛研究,目前已提出了大量的主动带宽测量算法.比较算法性能是设计精确、高效、健壮带宽测量算法的基础.本文定义算法的性能评价指标;通过大量带宽测量实验,从原理和实验结果分析指出误差累计和背景流量影响是影响现有测量算法性能的最主要原因.在此基础上,提出和实现一个任意链路带宽测量方法,该方法可以消除逐跳测量造成的误差累计和背景流量影响.

关键词: 带宽测量; 评价指标; 实验评价; 算法改进

中图分类号: TP393.06 **文献标识码:** A **文章编号:** 0372-2112 (2002) 12A-2142-04

Experimental Research on Bandwidth Measurement Technology and Method Improvement

XIE Gao-gang[1,3], TANG Yan-xia[2], ZHANG Da-fang[1], LI Zhong-cheng[3]

(1. *College of Computer and Communication, Hunan University, Changsha, Hunan* 410082, *China*;
2. *Graduate School, Xidian University, Xi'an, Shaanxi* 710071, *China*;
3. *Institute of Computing Technology, CAS, Beijing* 100080, *China*)

Abstract: Bandwidth metrics are necessary for network plan, management and performance optimization. For its merits, for example with capacity of measurement end-to-end bandwidth across several autonomous systems, a lot of algorithms on bandwidth active measurement were proposed. To compare and analyse existing methodologies is the basic for design of more precise, effective and robust bandwidth measurement algorithms. The metrics are proposed for performance evaluation of these methodologies. The limitations are discussed from principle and experimental result with bandwidth measurement experiment. Eliminating the affection of deviation accumulation and cross traffic is research direction for improving existing bandwidth measurement methodologies. An improved bandwidth measurement method is proposed for eliminating infections of deviation accumulation caused by hop by hop measurement and cross traffic.

Key words: bandwidth measurement; metrics for evaluation; experiment evaluation; method improvement

1 引言

随着 Internet 应用日益广泛,网络骨干基本上采用大容量的 DWDM/SDH 传输系统,骨干链路和接入链路带宽成倍增长,但网络应用性能依然是人们关注的主要问题.影响应用性能的原因包括端系统、应用与协议设计、网络容量、资源调度策略等,其中链路带宽是最重要的网络资源,是传输路径性能的主要决定因素之一.网络瓶颈与整体性能分析、容量规划以及通路有效带宽的测量必须依靠链路带宽的精确测量[1].优化带宽使用和分配是提升网络性能首要解决的问题.许多网络资源监控系统将网络带宽作为最重要的性能指标,进行全网范围的网络规划,减少网络瓶颈,分析提升网络性能.QoS 敏感应用必须了解业务通路端到端链路带宽(或有效带宽测试),并选择不同的服务质量或选择不同有效带宽的路由,因此带宽测试被广泛研究.

近来人们设计大量带宽测试算法和测试系统,主动测试方法[1~4]因其安全性和灵活性成为人们研究的热点,设计精确、高效、快速、健壮的主动带宽测试算法是带宽测试研究的目标.但如何通过精确指标评价算法的性能,并进行各种算法性能对比分析,一直没有相关研究.本文提出了带宽测试系统的性能评价指标,利用实验的方法评价了目前主要带宽测试算法,分析算法误差和造成误差的原因,并提出改进算法.

2 主动带宽测量技术与实现系统

IETF's IPPM 发布了不同的 RFC 定义了单向延迟、延迟抖动、丢包率等网络性能指标测试方法[5],但至今还没有一个 RFC 定义带宽测量.以下分析主动带宽测量技术的基本原理.

网络是路由器等网络节点通过链路进行相互连接的集合,各个节点通过逐跳转发数据包,实现信息从源端到目的端的传递.在一跳转发中,数据包通过输入链路到达路由器,路由器将该数据包放入缓存器的队列中,等待其他先到或者优先级更高的数据包被路由器转发.路由器由路由表查询算法将数据包转发到相应的输出端口,输出端口将数据包以数据位的形式发送到链路上,传输至下一个路由器的输入端口.由

收稿日期:2002-04-18;修回日期:2002-09-10
基金项目:国家 863 高技术发展计划项目(No.2002AA121032);国家自然科学基金资助(No.60273070, No.69973016)

以上的数据包转发过程描述,我们可以将数据包转发的延迟分为排队延迟(Queuing Time)、传输延迟(Transfers Delay)以及传播延迟(Transmission Latency)几个部分.假设端到端通路包含 n 跳,数据包 k 在第 i 跳排队等待时间为 $t_{iq}^{(k)}$,传输延迟为 $t_{id}^{(k)}$,传播时延为 $t_{il}^{(k)}$.路由器查找路由表以及数据包在路由器中其他的处理时间对特定的设备相对固定.假设链路 i 带宽 B_i(bps,bits/second),测试数据包 k 大小为 $S^{(k)}$ 比特,响应数据包大小为 $S^{(k')}$ 比特.显然数据包 k 在第 i 跳的延迟为

$$T_i^{(k)} = t_{iq}^{(k)} + t_{id}^{(k)} + t_{il}^{(k)} \tag{1}$$

则通路 P 总延迟为

$$T^{(k)} = \sum_{i\in P} T^{(k)}(i) = \sum_{i\in P} (t_{iq}^{(k)} + t_{id}^{(k)} + t_{il}^{(k)}) \tag{2}$$

对于特定的网络节点 $t_{id}^{(k)}$ 和 $t_{il}^{(k)}$ 是固定的,因背景流量的存在和流量的突发特性使得在缓存中排队的数据包数目不确定,即 $t_{iq}^{(k)}$ 是可变的,因此在测量端到端的延迟时,测量值具有较大的变化.

链路 l 的带宽可以表示为 $B_l = \max(P_{\Delta t})/\Delta t$,其中 $l\in L$,$\max(P_{\Delta t})$ 为任意 Δt 时间段内接口可以向传输介质(链路)发送的最大数据包的字节数据.

在 IP 网络中,链路带宽指在没有负载的情况下,链路可以提供给流的最大 IP 层吞吐量.由带宽定义知道,链路带宽表征接口处理数据包的能力,链路带宽越大,接口在单位时间内把数据包发送到传输介质上的能力就越强,数据包传送的速率就越快.链路带宽经常称为链路容量(Link Capacity).显然数据包 k 在链路 i 上的传输延迟为

$$t_{id}^{k} = S^{(k)}/B_i \tag{3}$$

即链路 i 总延迟为

$$T_i^{(k)} = t_{iq}^{(k)} + t_{id}^{(k)} + t_{il}^{(k)} = t_{iq}^{(k)} + S^{(k)}/B_i + t_{il}^{(k)} \tag{4}$$

链路带宽可以表示为

$$B_i = S^{(k)}/(T_i^{(k)} - t_{iq}^{(k)} - t_{il}^{(k)}) \tag{5}$$

根据测量的指标,可以将带宽测量分为端到端(End-to-end)带宽测量和链路(Hop-by-hop)带宽测量.端到端测量技术主要测量路径的容量和有效带宽[2~4];链路带宽测量主要测量路径中的各链路的容量和利用率[6~8].已知端到端链路径和各链路带宽,则可以知道路径带宽.

目前链路带宽主动测量技术基本上采用变包测量(VPS)法.VPS 最早由 Jacobson 实现[6],用以测量链路的带宽.假设已知 $S^{(k)}$,$t_{iq}^{(k)}$,$t_{il}^{(k)}$,$T(i)$ 则可以由式(5)测量链路容量.VPS 为消除 $t_{iq}^{(k)}$ 作如下假设:发送大小相同的探测数据包序列,假设延迟最小数据包没有在路由器中排队,即此时 $t_{iq}^{(k)} = 0$.VPS 测量带宽的方法可以描述为:向被测链路发送相同大小的探测包发送若干次,选取延迟最小的数据包,记录延迟和数据包的大小,回归分析计算链路带宽.

3 带宽测量系统性能比较实验

3.1 实验评价指标

设计健壮、高效、精确的带宽测量方法是带宽测量研究的核心内容.现有的带宽测量方法一般采用多次发送探测数据包,将 RTT 测量值进行统计回归分析得出带宽结论,因此往往需要持续较长的测试时间、发送较多的测试数据包.为了科学地评价测量方法和实现系统的优劣,我们定义了本次实验测量算法(实现系统)评价指标.

定义 1　测试持续时间 T_m:完成一次待测链路/路径带宽测试需要的时间.测试持续时间 T_m 主要由几部分构成:单次测试数据包发送和接收处理时间 T_p 以及统计回归处理时间 T_c,其中 T_p 由测试数据包发送时间 T_s,测试数据包传输时间 T_t 和测试数据包目的设备处理时间 T_d 组成.即

$$T = n * T_p + T_c = n * (T_s + T_t + T_d) + T_c \tag{6}$$

由式(6)知,测试持续时间主要由测量数据包发送次数决定.流量工程、QoS 路由等一般要求快速测定链路带宽,若测试持续时间越长,测试效率越低,测试负载对网络应用造成的影响就越大,而有效带宽又是动态变化的,则有效带宽测量值的实时性就越差.

定义 2　测量指标:测量方法可以完成链路带宽、路径带宽或者其他相关性能指标的测试.测量指标反映了该测量方法的测量能力.

定义 3　测量精度:假设测量得到的(有效)带宽值为 B_m,该链路或者通路实际带宽值为 B_0,则测试精度定义为

$$\eta = [(|B_m - B_0|)/B_0] * 100\% \tag{7}$$

B_0 一般取设计额定值(例如 SNMP MIB interface 组的参数值)或者理论值.

定义 4　测量稳定性 S:假设对目标链路进行 n 次测量,测量得到的带宽值分别为 $B_1, B_2, \cdots, B_n$,$\overline{B_i}$ 为平均值,则测量稳定性 S 定义为

$$S = [1 - (\sum_{i=1}^{n} |B_i - \overline{B_i}|)/(n * \overline{B_i})] * 100\% \tag{8}$$

测量精度和测量稳定性共同放映了测量算法的精确程度.

定义 5　测量负载:完成一次带宽测量需要发送和接收的数据包字节数,单位为 Byte.

测量负载越低则测量对网络运行的影响就越小.设计测量负载小、测量精度高、测量稳定性高、测量持续时间短的带宽测量方法是带宽测量技术研究的目标.

3.2 实验过程

本次测量实验中我们选择停业务测量和带业务测量两种方法进行实验,实验环境如图 1 所示.停业务测量中把 SmartBits2000 接入网络形成回路,逐步加大网络流量,直到产生丢包,适当减少流量,反复多次,选择不丢包情况下的最大流量值即为网络容量.停业务测量其实就是测量网络设备是否可以线速转发数据包.带业务测量中,测量软件安装在测量主机上,通过发送探测数据包完成链路带宽的测量.以下将对带业务测量进行详细的介绍分析.

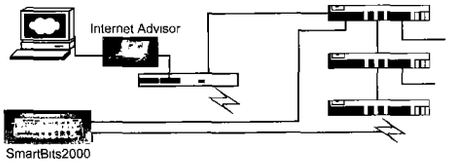


图 1　实验网络图

ICT 网络主要通过 CISCO2600 系列路由器实现各个研究室的互联,网络中的主机数大致为 1000 台,通过 10M 出口链路连接中国科技网.测试主机通过一 3COM SuperStack II 1100 接入网络.在链路中间串接 Agilent Internet Advisor 2300D 协议分析仪,串接端口为 2900LAN 分析基座提供的 100FE/E 端口.Agilent Internet Advisor 主要用于测量负载记录和数据包分析.将 Internet Advisor 设置在被动监听状态并设置过滤器,使得只有目标地址和源地址为测量主机地址的数据包被捕获.

测量主机的配置如下:(1)主机硬件配置:CPU(Intel Pentium III 667),Memory(128M),Hard Disk(20G),NIC(Network Interface Cards,3COM EtherLink XL 10/100 PCI TX NIC 3C905B-TX);(2)OS.:Red-hat 7.2 with a GNU/Linuxkernel 2.4.7-10,测量主机的 IP 为 159.226.39.204;(3)带宽测量工具包括 pathchar[6]、clink[7]、pchar[8].

3.3 实验数据

表 1 显示了 pathchar、clink 以及 pchar 实验基本数据.由表知 pchar 测量 3 跳需要的时间远远超过其他测量工具,但是三测量工具花费的时间最小需要也超过 10 分钟,因此使用该方法进行实时的负载均衡、QoS 路由以及流量工程是不现实的.测量带宽需要发送大量的探测数据包,即使测量负载最小的 clink 造成的测量负载也近 17Mb.

表 1 pathchar,clink 以及 pchar 实验基本数据

	Pathchar	Clink	pchar
测量组数	29	69	104
平均测量时间	18'32"	27'28"	52'45"
测量负载(Byte)	3317760	1705248	3320832
探测包平均速率(bps)	23868.8	8277.9	8393.9
带宽使用率(10Mbps)	0.22%	0.079%	0.080%

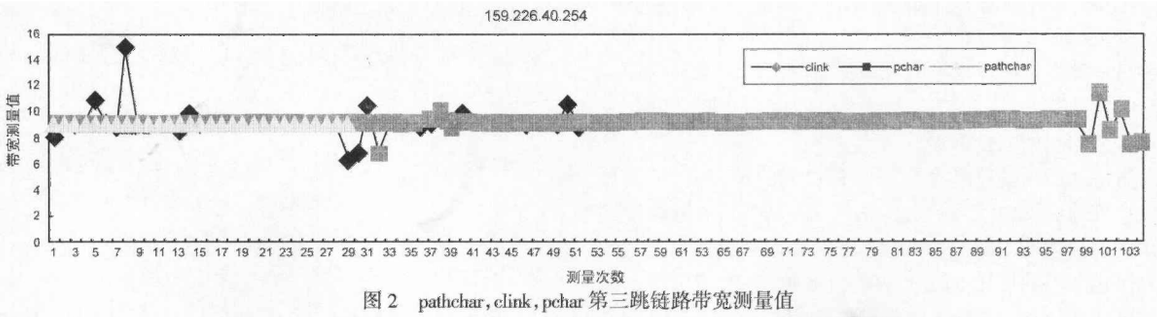


图 2 pathchar,clink,pchar 第三跳链路带宽测量值

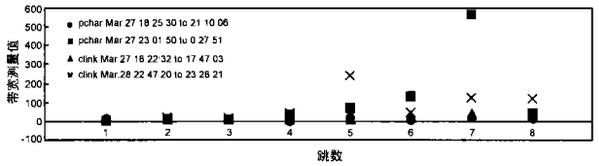
图 2 显示了第三跳链路带宽测量值.

以测量的数学期望作为带宽测量值,以停业务测量值作为实际带宽值,得到对各个测量方法的性能评价指标.表 2 表示 clink、pchar 和 pathchar 的三跳(159.226.40.254,159.226.41.62 和 1.1.1.1)各测量性能指标.

为了验证测量工具对多跳链路带宽测量有效性,我们对 159.226.39.204 至 www.sina.com.cn 进行逐跳的链路带宽测量.测量结果如图 3 所示.

表 2 测量算法性能指标

链路 / 统计指标	Pathchar			pchar			clink		
	1	2	3	1	2	3	1	2	3
带宽测量值(Mbps)	9.048	10.083	8.997	8.998	9.976	8.937	9.127	10.718	9.520
误差(%)	9.52	0.83	10.03	10.02	0.024	10.62	8.13	7.18	4.80
稳定性(%)	99.22	97.43	88.88	97.92	97.31	90.61	96.25	88.73	88.85



1:159.226.40.254;2:159.226.41.62;3:1.1.1.1;4:159.226.254.133;
5:159.226.254.146;6:202.97.15.17;7:202.97.37.185;8:202.96.12.42

图 3 多跳带宽测量带宽结果

4 实验结果分析

• 精确性 由图 3 和表 2 知道,随着跳数的增加,测量误差增加,精确性下降.但第二跳的精确性高于第一跳的精确性,且第一跳的测量值小于实际值,其原因在于测量主机通过两层设备接入网络,测量的延迟增大,测量值小于实际值.测量第二跳带宽时,不受第一跳的两层设备延迟影响,因此精度提高.逐跳测量累计误差的存在,使得跳数越多,精度越低.

• 稳定性 背景流量的存在和背景流量的突发性,使得测量数据包必然在各个路由器缓存中排队.排队的时间由背景流量大小决定,因而具有突发性,即测量在各个探测数据包时的延迟也具有突发性,估计的带宽具有随机性.背景流量突发性是造成测量值不稳定的直接原因.测量数据包需要在路径的各个路由器中排队.本测量实验采用的网络比较小,路由器间的连接关系简单,各个路由器缓存中的数据包数具有一定的时间相关性.随着跳数的增加,测量的延迟值变化越大,稳定性下降.

• 实时性 需要测量不同大小数据包情况下的延迟,并采用线性回归方法计算带宽,因此实验的各个工具都不适合实时测量,特别 pchar,测量三跳的时间达到 52 分钟.对于实时路由、流量工程等不适合使用该方法进行网络状态检测.

• 背景流量影响 背景流量直接影响 VPS 测量技术,背景流量越大,测量的延迟就越大,带宽测量值下降.但当链路 l 前面的链路带宽估计偏小时,某个探测包测量链路 l 带宽时恰好没有背景流量存在,则链路 l 延迟测量值偏小,测量带宽值比实际值大.测量带宽值变化,路由器缓存中数据包排队量,即反映了背景流量变化.例如在图 2 中,pchar 在横坐标 30 附近和 100 附近测量的带宽值变化比较大,而该测量时间为 14:00 至 19:00.由 SNMP MIB 采集的链路利用率知道,该时刻

对于的网络流量突发性最大.而横坐标 50 至 90 测量的时间为 2:00 至 6:00,背景流量比较小,因此测量带宽比较稳定.

• 多跳影响　图 3 所示,由于误差累计,在 4 跳以后带宽测量值随机性极大,三测量工具都不适合多跳链路的测量.第 5 小节中将提出一个改进的测量算法,消除背景流量影响.

• 测量负载影响　三测量工具需要发送和接收大量的测量数据包,因为测量的时间比较长,测量负载发送的平均强度并不大.

• 实验结论　各测量工具都是逐跳测量链路带宽,在跳数较少时具有一定的测量精度;测量值受背景流量影响非常大,在网络负载较大时,因为背景流量的突发性,造成网络测量值较大的随机性;因为是逐跳测量,需要较长的测量时间,不适合实时标定网络状态;因此各测量算法需要着重在测量精度、消除背景流量影响提高测量稳定性以及测量实时性展开研究.

5　算法改进

由以上实验分析知道,影响测量精度的主要缺陷在于背景流量和多跳的误差累计.本节我们提出测量改进方法,消除背景流量和多跳误差累计影响.

pathchar、clink 以及 pchar 进行逐跳(hop-by-hop)链路带宽测量,基于逐跳测量方法具有误差累积和实时性能差的缺点.接收端和发送端的时间同步问题,测量单向延迟远较测量双向延迟困难,因此我们往往测量双向延迟代替单向延迟.现有带宽测量算法假设探测数据包在路由器的缓存中是不排队的,或者假设在最小 RTT 时探测包在缓存中不排队,由实验测试知,背景流量的影响对测量精度有很大的影响.

假设发送 4 个数据包,将各个数据包分别按照发送顺序标识为 1,2,3,4,其中 $S^1 = S^2$, $S^3 = S^4$; $TTL^1 = TTL^3 = n$, $TTL^2 = TTL^4 = n-1$,且各数据包都是 back-to-back 包,即数据包发送的间隔相同 $T_1^2 = T_3^4$,并尽可能小,则由式(5)可以得到链路容量 B_n:

$$B_n = \frac{S^{(1)} + S^{(1')} - S^{(3)} - S^{(3')}}{RTT^1(n) - RTT^2(n-1) - RTT^3(n) + RTT^4(n-1)} \tag{9}$$

实现以上算法,进行在网络业务繁忙期间进行比较实验,得到的测量结果如表 3 所示.

表 3　业务繁忙期带宽测量值比较表

	测量时间	带宽测量值(Mbps)	测量时间	测量负载 Bytes
clink	2002-3-27,16:22:32-17:47:03	6.619	36'56"	—
pchar	2002-3-27,18:15:30-21:10:06	3.528	79'32"	—
改进算法	2002-3-27,10:45:45-10:49:22	9.230	5'37"	71200

6　结论

近年来 Internet 技术及其应用都有极大的改变,QoS 敏感的业务增加是 Internet 承载业务最大变化.提供 QoS 保证必然要求精确实时的性能测量技术,带宽测量是性能测量最重要的指标之一.本文定义了一系列带宽测量算法评价指标,对现有的带宽测量技术进行详细的实验和分析,通过实验结果分析知道影响 VPS 测量精度的最大问题在于背景流量和多跳误差累计.背景流量使得现有算法在链路利用率高时,测量结果具有较大的误差.多跳误差累计使得测量算法只适用于 4 跳以下的链路带宽测量.在此基础上提出消除背景流量和误差累积的方法,提出带宽测量的改进算法.

改进现有带宽测量算法的跳数和背景流量限制,并应用网络性能参数改进流控机制、动态网络资源管理、路由协议以及流量工程,提高整个网络的效率是未来带宽测量技术的主要研究方向.

参考文献:

[1] V Paxson. Measurements and analysis of end-to-end Internet dynamics [D]. USA: University of California, Berkeley, 1997.

[2] K Lai, M Baker. Measuring Link Bandwidths using a deterministic model of packet delay[A]. ACM SIGCOMM2000 Proceedings[C]. Sweden, 2000.283 - 294.

[3] K Lai, M Baker. Measuring bandwidth[A]. IEEE INFOCOM Proceedings[C]. New York, 1999.235 - 245.

[4] C Dovrolis, P Ramanathan, D Moore. What do packet dispersion techniques measure[A]? IEEE INFOCOM Proceedings[C]. Alaska, USA, 2001.905 - 914.

[5] RFC2330, V Paxson, G Almes, J Mahdavi, M Mathis. Framework for IP performance metrics[S]. May 1998.

[6] V Jacobson. Pathchar: a tool for measuring internet path characteristics [EB/OL]. ftp://ftp.ee.lbl.gov/pathchar/, April 1997.

[7] A B Downey. clink: a tool for estimation Internet link characteristics [EB/OL]. http://rocky.Wellesley.edu/downey/clink, June 1999.

[8] B A Mah. Pchar: a tool for measuring Internet path characteristics[EB/OL]. http://www.employees.org/~bmah/Software/pchar/, June 2000.

作者简介:

谢高岗　男,1974 年 5 月生于浙江衢州,2002 年获博士学位,现任中科院计算所副研究员,主要研究方向为网络测试与监控,服务质量.

汤艳霞　女,1976 年 2 月生于河南省济源市,西安电子科技大学 2000 级硕士生,中科院计算所客座研究生,主要研究方向为智能网络管理、网络测试与性能评价.

张大方　男,1959 年 4 月生于上海,获博士学位,现任湖南大学计算机与通信学院院长,教授,博士生导师.主要研究方向为容错计算与网络测试.

李忠诚　男,1962 年 11 月出生于黑龙江省哈尔滨,获博士学位,现任计算所研究员、博士生导师.主要研究方向为网络与通讯.