

网络非对称链路带宽的测量*

赵金** 陈鸣

(解放军理工大学通信工程学院 南京 210007)

【摘要】研究了数据分组在IP网络中传输时RTT时延的组成,给出了RTT时延与分组长度的线性表达式。分析了传统利用UDP测量分组RTT时延与测量分组长度的线性关系进行链路带宽测量的算法,针对该测量算法不能对非对称链路带宽进行测量的缺陷,提出了一种新的非对称链路带宽测量算法Asy_pathchar。同时使UDP报文和ICMP报文作为测量报文,Asy_pathchar能够对测量分组所经过的每一跳非对称链路的带宽进行测量,实验结果验证了算法的正确性。

关键词 链路带宽; 带宽测量; 非对称链路; RTT时延

中图分类号 TP393

Measuring Asymmetric Link Bandwidth

Zhao Jin Chen Ming

(Institute of Communications Engineering, PLA University of Science and Technology Nanjing 210007)

Abstract By studying the construction of the RTT latency, a linear expression between the RTT latency and the packet size is presented. Then pathchar, traditional bandwidth measurement algorithm, based on the relationship between the RTT latency and the packet size, is analyzed particularly. But pathchar can not measure the asymmetric link bandwidth. To solve this problem, Asy_pathchar, a new algorithm, is put forward, which can measure each hop's link bandwidth in both directions in an IP network using the UDP packet and ICMP packet. Finally an experiment is given to validate the new algorithm.

Key words link bandwidth; bandwidth measurement; asymmetric link; RTT latency

目前,比较流行的链路带宽测量算法pathchar及其衍生版clink和pchar都是通过向网络发送不同长度的测量分组^[1~4],在发送端测量该分组的RTT时延,根据分组长度与RTT时延变化规律来推算出分组所通过链路的链路带宽。虽然它们能够比较精确地测量出从源到目的的路径上每一段链路的链路带宽,但都假定链路对称,而ADSL、卫星链路、56 K调制解调器等技术的运用,已经改变了这种假设,这对非对称链路带宽的测量、发现链路的瓶颈带宽、提高和改善网络的性能等方面都有重要意义。

1 测量网络模型与RTT时延的组成

基于发送端的带宽测量算法所使用的测量分组一般都是UDP报文或ICMP echo报文。采用UDP报文,应答报文都是长度为56字节的ICMP超时报文;采用ICMP echo报文,应答报文是与测量报文

2002年2月5日收稿

* 国家863高科技项目

** 男 30岁 博士

长度相同的ICMP reply报文或长度为56字节的ICMP超时报文。链路带宽测量算法Pathchar所用的测量分组是UDP报文。本文提出的非对称链路带宽测量算法使用的是ICMP echo报文,其网络模型如图1所示,下面讨论测量分组从节点 $(n-1) \sim n$ 的RTT时延的组成。

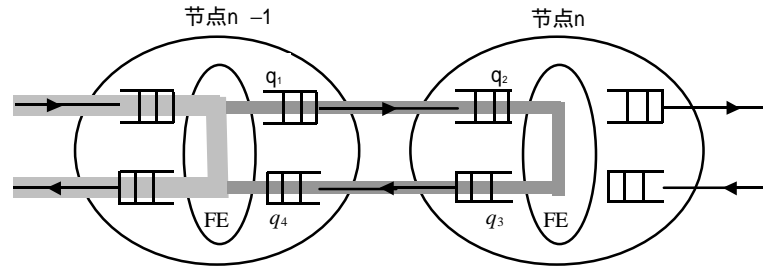


图1 网络模型

设测量分组的长度为 S bit,从节点 $(n-1) \sim n$ 的链路 l_n 的链路带宽为 C_n bit/s,从节点 $n \sim (n-1)$ 的链路 l'_n 的链路带宽为 C'_n bit/s。测量分组在进入链路 l_n 之前在节点 $(n-1)$ 中的排队时延为 q_1 s,进入链路 l_n 时的发送时延为 S/C_n ,在链路 l_n 上的传播时延为 d_n s。进入节点 n 后的排队时延为 q_2 s。节点 n 对测量分组进行处理,这时会有两种情况发生:1) TTL值减1后为0,则产生长度为56字节的ICMP超时报文;2)若测量分组为ICMP echo报文,则产生同样长度的ICMP reply报文;设节点对测量分组的处理时间为 t_p s,响应报文的长度为 S' bit。响应报文从节点 n 进入链路 l'_n 以前,在发送队列中的排队时延为 q_3 s,进入链路 l'_n 时的发送时延为 S'/C'_n ,响应报文在链路 l'_n 上的传播时延为 d'_n s,到达节点 $(n-1)$ 后等待处理时的排队时延为 q_4 s。所以,测量分组从节点 $(n-1) \sim n$ 的RTT时延为

$$\text{RTT} = q_1 + \left(\frac{S}{C_n} + d_n\right) + q_2 + t_p + q_3 + \left(\frac{S'}{C'_n} + d'_n\right) + q_4 \quad (1)$$

2 Pathchar链路带宽测量算法

文献[1]提出Pathchar测量算法,它是目前网络链路带宽测量算法中最成功的算法之一,是唯一的一种只需在一个终端主机中安装一些测量软件就能够比较精确地测量出一条路径中各段链路带宽的算法。

Pathchar在测量第 n 段链路的链路带宽时,在发送端向接收端发送一系列长度不同的TTL值为 n 的UDP测量分组,然后在发送端测量每一个测量分组超时报文返回的时刻,由此计算出每一个测量分组的RTT时延。发送测量分组时,必须在收到前一个测量分组的超时报文后,才能发送下一个测量分组。通过对测量分组长度与RTT时延的统计分析,就可以得出测量分组长度、RTT时延和链路带宽之间的关系。

由式(1)可知,测量分组从节点 $(n-1) \sim n$ 的RTT时延为

$$\text{RTT} = q_1 + \left(\frac{S}{C_n} + d_n\right) + q_2 + t_p + q_3 + \left(\frac{56 \times 8}{C'_n} + d'_n\right) + q_4$$

为了简化上式,Pathchar作了以下两个假设:1) 处理机处理的时间足够快,处理时延 t_p 可以忽略不计;2) 测量时被测链路上无其他的流存在,即忽略排队时延。所以,上式化简为

$$\text{RTT} = \frac{S}{C_n} + \frac{56 \times 8}{C'_n} + d_n + d'_n$$

在实际测量时,是从发送端测量第 n 段链路带宽,此时RTT时延 T_n 应为

$$T_n = \sum_{i=1}^n \left(\frac{S}{C_i} + \frac{56 \times 8}{C_i} + d_i + d_i' \right) = S \sum_{i=1}^n \frac{1}{C_i} + \sum_{i=1}^n \left(\frac{56 \times 8}{C_i} + d_i + d_i' \right)$$

式中 对于每个确定的 n ，不同的 S 值都有相应的 T_n 值与其对应，可视 T_n 为 S 的线性函数，设 k_n 表示线性函数 $T_n(S)$ 的斜率，则

$$k_n = \sum_{i=1}^n \frac{1}{C_i} \Rightarrow k_n - k_{n-1} = \frac{1}{C_n}$$

式中 $n=0$ 时， $k_1 = \frac{1}{C_1}$ 。所以

$$C_n = \frac{1}{k_n - k_{n-1}} \quad (2)$$

从式(2)可知，要想得到 C_n ，首先必须求得 k_n 与 k_{n-1} 。为得到 k_n ，需要不断地向网络发送长度不同的测量分组，在发送端测量每个分组的RTT时延 T_n 。就得到一系列的样本点 (S, T_n) ，通过 S 与 T_n 的线性关系，就可以根据这些样本点绘出一条直线，直线的斜率即所求的 k_n ，同理可得 k_{n-1} ，从而可进一步算出被测链路的链路带宽 C_n 。

通过对Pathchar测量原理的分析，可以看到，不论是否假定链路的对称性，Pathchar算法只能对从源到目的单方向路径上各段链路的带宽进行测量，而不能对反方向路径上各段链路带宽进行测量。针对Pathchar的这一缺陷，本文提出了一种能够测量非对称链路带宽的测量算法Asy_pathchar。

3 Asy_pathchar非对称链路带宽测量算法

在Asy_pathchar带宽测量算法中，运用Pathchar算法的基本原理。所不同的是在Asy_pathchar算法中，所用的测量分组不再是UDP报文，而是ICMP echo报文。本文的目的是在节点 n 处，对同一长度的测量分组，产生两种类型的应答报文。

在测量第 n 跳链路的带宽时，首先向节点 n 发送TTL= n 的ICMP echo报文，在节点 n 处，产生与测量分组长度相同的ICMP reply响应报文；发送端收到ICMP reply报文的返回后，再向节点 $n+1$ 发送同样长度TTL= n 的ICMP echo报文；在节点 n 处，TTL值减为0，但该报文的目的是节点 n 的下一个节点 $n+1$ ，所以节点 n 判断该报文超时，向发送端返回长度为56字节的超时报文。在发送端分别计算出两次发送ICMP echo报文的RTT时延。通过对两次RTT时延组成的分析，就可以计算出上行链路和下行链路的带宽。

如响应报文为ICMP超时报文，由式(1)可知，测量分组从节点 $(n-1) \sim n$ 的RTT时延为

$$\text{RTT} = q_1 + \left(\frac{S}{C_n} + d_n \right) + q_2 + t_p + q_3 + \left(\frac{56 \times 8}{C_n} + d_n' \right) + q_4$$

与Pathchar中的结果完全相同，因此可得到与Pathchar中同样的结论

$$C_n = \frac{1}{k_n - k_{n-1}} \quad (3)$$

如响应报文为ICMP reply报文，由式(1)可知，测量分组从节点 $(n-1) \sim n$ 的RTT时延为

$$\text{RTT} = q_1 + \left(\frac{S}{C_n} + d_n \right) + q_2 + t_p + q_3 + \left(\frac{S}{C_n} + d_n' \right) + q_4$$

采用与Pathchar中同样的假设，上式化简为

$$\text{RTT} = \frac{S}{C_n} + \frac{S}{C_n} + d_n + d_n'$$

在实际测量时，即从发送端测量第 n 段链路带宽，此时RTT时延 T_n 应为

$$T_n = \sum_{i=1}^n \left(\frac{S}{C_i} + \frac{S}{C_i} + d_i + d_i' \right) = S \sum_{i=1}^n \left(\frac{1}{C_i} + \frac{1}{C_i} \right) + \sum_{i=1}^n (d_i + d_i')$$

式中 对于每个确定的 n , 不同的 S 值都有相应的 T_n 值与其对应, 可视 T_n 为 S 的线性函数, 设 k_n' 表示线性函数 $T_n(S)$ 的斜率, 则

$$k_n' = \sum_{i=1}^n \left(\frac{1}{C_i} + \frac{1}{C_i} \right) \Rightarrow k_n' - k_{n-1}' = \frac{1}{C_n} + \frac{1}{C_n} \quad (4)$$

式中 $n=0$ 时

$$k_1' = \frac{1}{C_1} + \frac{1}{C_1}$$

将式(3)的结果代入式(4), 可得

$$C_n' = \frac{1}{(k_n' - k_{n-1}') - (k_n - k_{n-1})} \quad (5)$$

因此, 通过式(3)和式(5), 就可以测量出非对称链路上行链路带宽和下行链路带宽。

4 实验验证

本文设计了一个实验来验证Asy_pathchar算法的正确性, 其实验原理如图2所示。

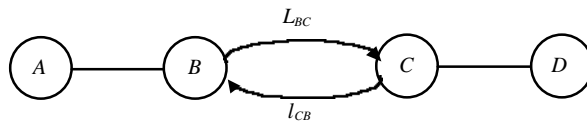


图2 实验原理图

图2中4台机器均为P 500的微机, 其中B与C为插有1块网卡, 2块modem的多穴主机, 2块modem的速度为56 Kb/s和14.4 Kb/s。链路 l_{BC} 的带宽为56 Kb/s, 链路 l_{CB} 的带宽为14.4 Kb/s。在机器B和C中精心设计路由表, 使从 $B \sim C$ 和从 $C \sim B$ 方向的数据转发分别经过不同的链路。

首先采用Pathchar带宽测量法分别在机器A和机器D处测出链路 l_{BC} 和链路 l_{CB} 的链路带宽, 然后采用Asy_pathchar测量法在机器A处测量链路 l_{BC} 和链路 l_{CB} 的链路带宽, 再对两种测量方法的测量结果进行比较, 测得的结果一致, 从而证明了Asy_pathchar算法能够在发送端测量出非对称链路所链路带宽, 其测量结果如表1所示。

表1 两种测量算法测量结果比较

测量方法	l_{BC} 的链路带宽/ $\text{Kb} \cdot \text{s}^{-1}$	l_{CB} 的链路带宽/ $\text{Kb} \cdot \text{s}^{-1}$
Pathchar	55.8	14.1
Asy_pathchar	55.8	14.1

5 结束语

本文提出的Asy_pathchar非对称链路带宽测量算法, 继承了Pathchar算法的优点, 即只需在一个终端主机中安装一些测量软件就能够比较精确地测量出端到端的路径中各段链路的带宽, 同时又克服了Pathchar不能测量非对称链路带宽的缺陷。实验证明, Asy_pathchar对非对称链路带宽测量是行之有效的。但在实际使用中要注意一个问题, 如果节点路由器屏蔽了ICMP echo报文, Asy_pathchar就不能正常的工作。

参 考 文 献

- 1 Jacobson V. Pathchar-A tool to infer characteristics of internet paths. <ftp://ftp.ee.lbl.gov/pathchar/>, April 1997
- 2 Downey A. Clink-A tool for estimating internet link characteristics. <http://rocky.wellesley.edu/downey/clink/>, June 1999
- 3 Mah B A. Pchar-A tool for measuring internet path characteristics. <http://www.employees.org/bmah/Software/pchar/>, June 2000
- 4 Downey A. Using pathchar to estimate internet link characteristics. ACM SIGCOMM, 1999: 241~250
- 5 Paxson V. Measurements and analysis of end-to-end internet dynamics: [PhD thesis]. Berkeley: niversity of California, 1997

· 科研成果介绍 ·

复合压电材料研究

主研人员：张树人 傅应泉 郭曙光 钟朝位 周晓华 蔡雪梅 余昌奎 黄富钊 韩冬 杜广山

在复合压电材料用陶瓷基阵的组成及性能的优化研究中，采用先进的原料处理及致密化烧结技术，通过改进高温极化工艺，制备了性能优良的收发兼用的PZT陶瓷材料。通过对1-3型和2-0-2型复合材料的结构设计与加工工艺研究，解决了精细切割、陶瓷聚合物匹配相容等关键技术，尤其是采用横向切割-填充-纵向切割-填充工艺，解决了精细切割中磁体碎裂的技术难题。通过对复合材料结构参数的优化，制备了性能优良的1-3型和2-0-2型复合压电材料。

互连网信息记录与审计系统

主研人员：汪文勇 黄鹂声 杨红 蒋锡民 程圣宇

互连网信息记录与审计系统能捕获以太网帧并识别IP分组，对UDP包进行日志记录，对TCP数据流进行记录、重组和还原；对HTTP，FTP、SMTP、POP3、TENET等应用协议进行解析和数据处理；对日志进行查询和分析。该系统提供了中继方式和旁听方式，并采用C/S模式，实现了对高层协议的记录与分析功能，具有独创性。该系统还提供了互连网络安全管理的一种有效手段，能针对可能发生的内部泄密和来自外部的攻击行为进行记录，并作为安全审计的依据。该系统具有较高的技术起点，运行稳定，功能较全面。

· 科 卞 ·