

# 基于四元组的链路双向带宽测量方法

贾圣文, 高仲合, 赵金龙

(曲阜师范大学计算机科学学院, 山东 日照 276826)

**摘要:** 为解决网络带宽测量中链路不对称的问题, 提出一种能在 IP 网络中测量出每条链路的双方向带宽的方法。采用由 TCP 报文和 ICMP 报文组成的四元组作为探测报文, 通过分析各个分组在 IP 网络中的 RTT 时延组成, 并利用分组大小差值与其往返时延差值的线性关系计算出链路的双向带宽。仿真实验证明, 该方法能快速准确地测量出任意链路的双向带宽。

**关键词:** 四元组; 双向带宽; 阈值; IP 网络

## Measure Approach of Link Bi-directional Bandwidth Based on Tetrad

JIA Sheng-wen, GAO Zhong-he, ZHAO Jin-long

(College of Computer Sciences, Qufu Normal University, Rizhao 276826, China)

**【Abstract】** To solve the asymmetric problem in network bandwidth measurement, this paper puts forward an approach based on a tetrad composed of TCP and ICMP packets to measure the bi-directional bandwidth of each link in an IP network. It calculates the bi-directional bandwidth by analyzing the composition of RTT latency of each packet in an IP network and using the linear relationship between the size difference of packets and that of RTT latencies. Both theoretical analysis and simulation experiment are employed to validate the new approach.

**【Key words】** tetrad; bi-directional bandwidth; threshold value; IP network

DOI: 10.3969/j.issn.1000-3428.2011.14.029

### 1 概述

准确的网络带宽对网络设计者、管理者和使用者都有重要意义, 因此, 如何对网络带宽进行准确的测量是迫切需要解决的问题。现有的大部分测量方法都假定链路是对称的, 但是 ADSL、cable modem、卫星链路都存在链路非对称现象, 无线网络中也存在大量非对称现象。非对称链路带宽测量方法按测量过程是否需要节点路由器的协作, 可分为链路带宽测量和端到端的路径带宽测量。文献[1]属于路径带宽测量, 旨在测量出路径中的前向、后向瓶颈带宽。文献[2-4]属于链路带宽测量, 旨在测量出每跳链路的前向、反向带宽。目前对非对称链路带宽的测量都是基于包间隔模型<sup>[5]</sup>的。

本文提出的基于四元组的链路双向带宽测量方法属于链路级带宽测量, 它采用的探测报文是一个由 TCP 报文和 ICMP 报文组成的四元组。

### 2 网络链路带宽测量模型

传统的链路带宽测量方法的基本原理是发送端向目的端发送不同长度的报文, 然后根据报文长度与 RTT 时延的线性关系来推算出链路的带宽。目前通用的链路带宽测量模型如图 1 所示。

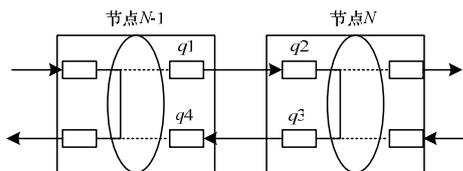


图 1 链路带宽测量网络模型

图 1 表示从节点 N-1 到节点 N 的 RTT 时延构成测量分组长度为 S bit, 节点 N-1 到节点 N 的带宽是  $C_n$  bit/s, 节点 N 到节点 N-1 的带宽是  $C_n'$  bit/s, 测量分组进入  $L_n$  链路之前在节点 N-1 排队的时间为  $q_1$ , 进入链路  $L_n$  的发送时延是  $S/C_n$ ,

在链路  $L_n$  的传播时延是  $d_n$ , 进入节点 N 后的排队时延是  $q_2$ , 节点 N 收到测量分组后进行处理, 产生长度为 S' 的应答报文, 节点对测量分组的处理时间为  $te$  s, 响应报文进入链路  $L_n'$  之前在发送队列时间为  $q_3$ , 进入  $L_n'$  发送时延为  $S'/C_n'$ , 响应报文在  $L_n'$  的传播时延为  $d_n'$ , 到达节点 N-1 后等待处理的时间为  $q_4$ 。对于以上的设定则测量分组从节点 N-1 到节点 N 的 RTT 时延为:

$$rtt = q_1 + (S/C_n + d_n) + q_2 + te + q_3 + (S'/C_n' + d_n') + q_4 \quad (1)$$

### 3 四元组非对称链路带宽测量方法

文献[2-3]的测量方法基于变长报文测量, 利用 ICMP 2 种不同类型的应答报文测量双向带宽。这种测量为了使排队时延忽略不计, 每种不同大小的测量包必须发送多次, 选择 RTT 最小的那一组。这样测量花费的时间周期长、测量速度慢、损耗的网络带宽大, 特别是在网络负载较大时, 很难得到无排队时延的分组, 这时测得的结果受到很大影响。同时在计算链路的反向带宽时要受前向带宽测量结果的影响, 存在累积误差。另外, 在实际测量过程中, 只能精确地得知测量分组在网络层的长度, 但是在测量过程中得到的 RTT 往返时延却不是长度为 s 的分组的 RTT 时延, 而是测量分组经过数据链路层和物理层后加上的额外字节  $\Delta$ , 即长度为  $(s+\Delta)$  的分组的往返时延, 这样就给测量结果带来了不小的误差, 使得测量结果偏小。文献[4]提出的 Asy\_pro 测量方法采用一次发送一对报文, 提高了测量性能, 但是仍然存在累积误差。

#### 3.1 探测报文的设计

为了解决上述不足, 本文提出了一种基于四元组的非对称链路带宽测量方法, 该方法中的测量分组是由 2 对报文组

**作者简介:** 贾圣文(1986—), 男, 硕士, 主研方向: 网络测量; 高仲合, 教授; 赵金龙, 硕士

**收稿日期:** 2011-02-17 **E-mail:** j\_swen@yahoo.cn

成的一个四元组，每对报文分别由一个 TCP 报文和一个 ICMP 报文组成，形如{[pq][rm]}，其中，每对报文对背靠背发送，2 对报文发送的时间间隔也尽可能小，报文  $p$ 、 $r$  是 TCP 报文， $q$ 、 $m$  是 ICMP 报文。在这个四元组中，包的大小为  $S(p) = S(q) = S(r) = S1$ ， $S(m) = S2$ ， $S1 \neq S2$ ，目的地址分别为  $n, n-1, \dots, 1$ 。这样当 4 个报文到达目的端时，无论是 TCP 报文还是 ICMP 报文都向源端发送长度分别为  $S_{ack}$  和  $S_{reply}$  的 ack 确认信息和 ICMP-reply 应答信息。

3.2 测量过程

按照上文设计好的探测报文，其具体的测量过程如下：

(1)源端发送探测报文。源端向目的端  $n$  发送如上述设计好的探测报文串{[pq][rm]}。

(2)接收报文并响应。目的端收到探测报文后分别向源端发送应答报文，对于探测报文  $p$ 、 $r$ ，目的端向源端发送大小为  $S_{ack}$  的 ack 确认信息；对于探测报文  $q$ 、 $m$ ，目的端向源端发送大小为  $S_{reply}$  的 ICMP-reply 报文。

(3)源端统计往返时延。源端收到应答报文后分别统计 4 个探测报文的往返时延。

(4)作差值计算并过滤。根据统计到的往返时延分别计算探测报文  $q$ 、 $p$  的时延差值  $\Delta T_{n\_qp}$  和探测报文  $m$ 、 $q$  的时延差值  $\Delta T_{n\_mq}$ 。根据报文  $p$  和报文  $r$  的往返时延差值大小过滤掉受背景流量影响较大的值，因为 2 个报文是完全相同的，如果探测过程中受背景流量影响不大的话，它们的往返时延也应该近似相等，只有那些  $rtt_r - rtt_p < \alpha$  的值才是符合条件的，否则丢弃。

(5)改变目的端重新发送。修改目的端分别为  $n-1, n-2, \dots, 1$ ，重复上述过程，得到源端到路径上任意一条链路的往返时延。

(6)计算每跳链路的双向带宽。按照如下的分析计算各链路的双向带宽。

由式(1)可得出四元组在链路  $L_{n-1}$  和链路  $L_n$  之间的往返时延分别是：

$$rtt_p = q_{1p} + (S1/C_n + d_{np}) + q_{2p} + t_{ep} + q_{3p} + (S_{ack}/C_n + d_{np}') + q_{4p} \tag{2}$$

$$rtt_q = q_{1q} + (S1/C_n + d_{nq}) + q_{2q} + t_{eq} + q_{3q} + (S_{reply}/C_n + d_{nq}') + q_{4q} \tag{3}$$

$$rtt_r = q_{1r} + (S1/C_n + d_{nr}) + q_{2r} + t_{er} + q_{3r} + (S_{ack}/C_n + d_{nr}') + q_{4r} \tag{4}$$

$$rtt_m = q_{1m} + (S2/C_n + d_{nm}) + q_{2m} + t_{em} + q_{3m} + (S_{reply}/C_n + d_{nm}') + q_{4m} \tag{5}$$

各符号与式(1)的符号表示的意义相同。

令  $\Delta rtt1 = rtt_q - rtt_p$ ， $\Delta rtt2 = rtt_m - rtt_q$ ， $\Delta S = S1 - S2$ 。根据算法的特点作以下假设：(1)由于对等机处理速度特别快，加上  $S1$  和  $S2$  相差不大，假设  $t_{ep}$ 、 $t_{eq}$ 、 $t_{er}$ 、 $t_{em}$  近似相等，在计算  $\Delta rtt1$ 、 $\Delta rtt2$  时可以相互抵消。(2)由于在测量过程中是发送分组对，发送间隔很小，当某段时间内网络流量稳定时各分组的排队时延近似相等，在计算  $\Delta rtt1$ 、 $\Delta rtt2$  时可相互抵消。

链路的传播时延仅与链路的物理特性有关，在计算  $\Delta rtt1$ 、 $\Delta rtt2$  时当然可以抵消，因此，得到  $\Delta rtt1 = (S_{reply} - S_{ack})/C_n$ ， $\Delta rtt2 = \Delta S/C_n$ 。

那么，从发送端测量第  $N$  跳的往返时延差值分别为：

$$\Delta T_{n\_qp} = \sum_{i=1}^n (S_{reply} - S_{ack}) / C_i \tag{6}$$

$$\Delta T_{n\_mq} = \sum_{i=1}^n \Delta S / C_i \tag{7}$$

则：

$$\Delta T_{n\_qp} - \Delta T_{n-1\_qp} = (S_{reply} - S_{ack}) / C_n'$$

即：

$$C_n' = (S_{reply} - S_{ack}) / (\Delta T_{n\_qp} - \Delta T_{n-1\_qp}) \tag{8}$$

$$\Delta T_{n\_mq} - \Delta T_{n-1\_mq} = \Delta S / C_n$$

即：

$$C_n = \Delta S / (\Delta T_{n\_mq} - \Delta T_{n-1\_mq}) \tag{9}$$

当  $n=1$  时， $\Delta T_{0\_qp} = \Delta T_{0\_mq} = 0$ 。

(7)改变探测报文的大小，重复计算。

这样就测到了路径中每条链路的双向带宽。

3.3 算法分析

本算法由于采用时延差值计算，可以不必假设分组的排队时延为 0，而只须假设各分组的排队时延近似相等，由于分组长度相差不大，发送间隔较小，从而在网络流量比较稳定的情况下相对比较容易保证这点。而无须像文献[2-3]那样为得到最小往返时延每种测量包必须发送多次。这样相对节省了大量的测量时间，加快了测量速度，测量损耗的带宽也相对较小，而且在本算法中采用一个阈值  $\alpha$  过滤掉那些 2 个分组对所受背景流量影响相差大的值。

另外，目前的各类工具都忽略各种链路层协议带来的误差影响，但不管是 Ethernet 或者 ATM 网络、RPR(弹性分组环)或者 SDH 光网络，还是无线网络，链路层协议都会占用一些物理带宽开销，其带来的误差影响是不能忽略的。本文采用包长度之差计算，消除了包经过数据链路层和物理层额外增加的长度的影响。假设增加的额外长度为  $\Delta$ ，采用包对长度之差： $S1 - S2 = (S1 + \Delta) - (S2 + \Delta)$ ，同理  $S_{reply} - S_{ack} = (S_{reply} + \Delta) - (S_{ack} + \Delta)$ ，这样通过这个相减关系，链路层报头长度  $\Delta$  就被抵消了。

4 实验仿真

分别在 2 种场景(图 2、图 3)下，用 ns2 仿真。在下列仿真结果图中用 Bw1、Bw2、Bw3 分别表示链路 1、链路 2、链路 3 的前向带宽值，Bw1'、Bw2'、Bw3' 分别表示链路 1、链路 2、链路 3 的反向带宽值。实验中所用阈值  $\alpha$  取值为  $1 \mu s$ 。

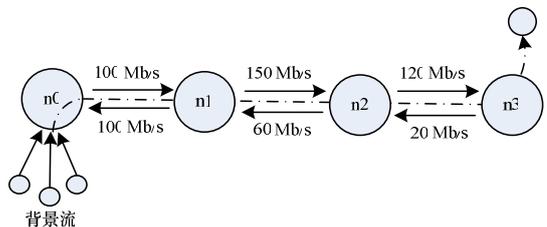


图 2 场景 1 实验拓扑图

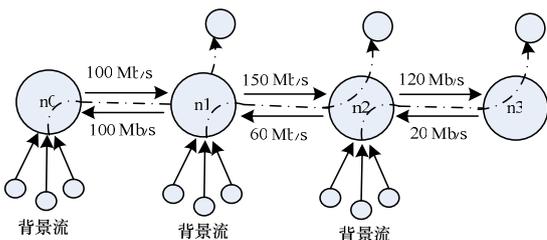


图 3 场景 2 实验拓扑图

4.1 场景 1 仿真实验

背景流量由指数分布流量产生器产生，速率为 60 Mb/s，

“ON”、“OFF”状态时间为200 ms, 发送60个大小不同的探测报文组, 用阈值过滤掉不符合条件的值, 仿真结果如图4和图5所示。

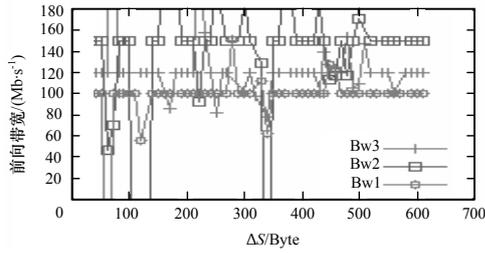


图4 Exponential 背景流量下的前向带宽

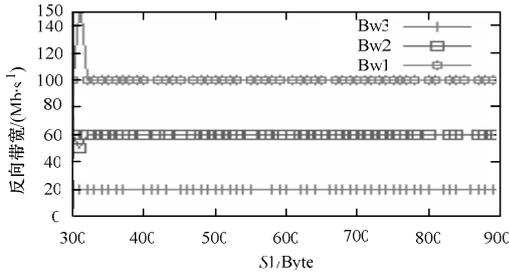


图5 Exponential 背景流量下的反向带宽

实验表明在这种路径背景流量下, 背景流量对测量结果影响不大, 用阈值  $\alpha$  过滤掉少量受背景流量影响较大的值后, 测量结果能准确反映出每跳链路的双向带宽。

### 4.2 场景2 仿真实验

图3是在每跳背景流量环境下的实验拓扑图, 其中, 背景流量由泊松流量产生器产生, “ON”、“OFF”状态时间为200 ms, “shape”为1.2, 发送60个大小不同的探测报文组, 用阈值过滤掉不符合条件的值, 图6和图7是在链路利用率40%下的仿真结果。

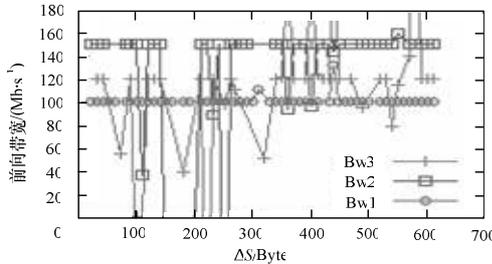


图6 每跳背景流量下的前向带宽

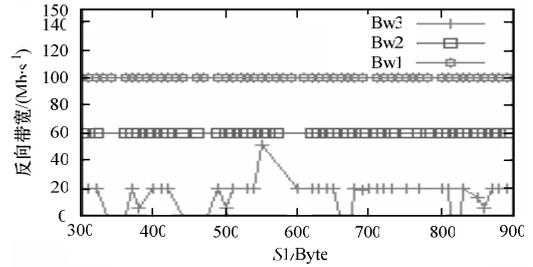


图7 每跳背景流量下的反向带宽

实验表明, 在每跳背景流量下的测量结果所受影响较路径背景流量环境下较大, 用阈值  $\alpha$  过滤掉受影响较大的值后仍能准确反映出每跳链路的双向带宽。

### 5 结束语

本文提出基于四元组的链路双向带宽测量方法。采用阈值过滤, 避免了VPS模型中每种大小包发送若干次的问题, 节省了测量时间, 加快了测量速度, 测量损耗的带宽也相对较小; 采用减法模型, 消除了包经过数据链路层和物理层额外增加长度的影响, 提高了测量精度。总之, 该方法能以较小的测量代价、高精度地测量出链路的双向带宽。

### 参考文献

- [1] Chen L J, Sun T, Yang Guang, et al. End-to-end Asymmetric Link Capacity Estimation[C]//Proc. of IFIP Networking'05. Waterloo, Canada: [s. n.], 2005.
- [2] Jiang Wenyu. Detecting and Measuring Asymmetric Links in an IP Network[C]//Proc. of IEEE Global Communications Conference. Rio de Janeiro, Brazil: [s. n.], 1999: 1727-1731.
- [3] 赵金, 陈鸣. 网络非对称链路带宽的测量[J]. 电子科技大学学报, 2002, 31(4): 404-408.
- [4] Wang Rong, Cheng Aiwang, Du Wei, et al. An Improved Asymmetric Link Bandwidth Measurement Algorithm[C]//Proc. of International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing. Shanghai, China: [s. n.], 2009: 608-611.
- [5] 于靖, 杨英杰, 常德显. 可用带宽测量算法的研究[J]. 计算机工程, 2009, 35(6): 136-138.

编辑 任吉慧

(上接第90页)

通过仿真实验的验证, 该异常概率能较合理地反映缺失数据的异常可能性。实际中可以通过缺失数据的异常概率直接找出某些肯定为异常的信号, 并根据实际情况对剩余的缺失数据按异常概率的大小进行选择性的处理。该方法一定程度上减小了后续数据处理的工作量, 并提高了处理效率, 具有一定的应用价值。

### 参考文献

- [1] Han J, Kamber M. Data Mining Concepts and Techniques[M]. San Francisco, USA: Morgan Kaufmann Publishers, 2006.
- [2] 廖学清, 吕强, 单冬冬. 数据缺失下学习贝叶斯网的SEM算法[J]. 计算机工程, 2009, 35(8): 214-216.

编辑 索书志

- [3] Li Jingrong, Li Pheng, Shu Khoo. RMINE: A Rough Set Based Data Mining Prototype for the Reasoning of Incomplete Data in Condition-based Fault Diagnosis[J]. Journal of Intelligent Manufacturing, 2006, 17(1): 163-176.
- [4] Qin Yongsong, Zhang Shichao, Zhu Xiaofeng. POP Algorithm: Kernel-based Imputation to Treat Missing Data in Knowledge Discovery from Databases[J]. Expert Systems with Application, 2009, 36(2): 2794-2804.
- [5] Little R, Rubin D. Statistical Analysis with Missing Data[M]. New York, USA: John Wiley & Sons, 2002.



