基于网络报文对的网络瓶颈带宽测试技术的研究

〖摘要〗 本文讨论了一种基于网络报文对的网络瓶颈带宽测试技术的原理及实现时一些问题的解决方案。给出了各种典型网络环境下的测试结果，并讨论了这种测试方法的潜在误差。

关键词 网络性能管理(Network Performance Management) 带宽测试(Bandwidth Measurement) 报文对(Packet Pair)

1. 引言

网络中源点与宿点的通路带宽取决于这条通路上速率最慢的信道的带宽。速率最慢的信道称之为瓶颈信道(Bottleneck Link)，其带宽相应称之为瓶颈带宽(Bottleneck Bandwidth)。网络瓶颈带宽的发现与测量对于管理和提高通路带宽，改善网络性能提供了重要信息。这里所讨论的带宽是指在没有任何竞争流量的情形下,所能获得的最大带宽，即由物理信道的性质所决定的基准带宽(Base Bandwidth)。传统的 ping 或 traceroute 等性能测试工具获得的是通路的传输延迟值，从而使用户直觉地感受到获得服务所需的时间概念。但是也有许多用户感兴趣的是网络的通过能力，即吞吐量，例如当他准备启动 FTP 服务来下载一个大数据之前。这个指标不能简单地从传输延迟中推断或换算出来，因此需要使用另外的方法来测量指定网络通路的通过能力。本文给出了一种利用报文对原理进行网络瓶颈带宽测试的方法。

它是在波士顿大学 Robert L. Carter 及 Mark E. Crovella 所提出方法的基础上进行了改进而得，改进后的方法不仅可以测出通路的瓶颈带宽而且可以准确的定位出瓶颈信道的位置。它的基本原理是根据观察到的两个相邻报文在通路上的离去时间差，求出瓶颈信道的带宽并可相应确定瓶颈信道的位置。

2. 报文对方法测试瓶颈带宽的原理分析

假设两点之间的通路由 n 跳的信道构成，长度为 P 的报文在带宽为 B 的信道上的延迟为 $D = P / B$。用 B_b 标识瓶颈信道带宽。测试原理如下:

测试主机在通路一端向通路的另一端在极短的时间间隔 T 内接连发送出两个报文长度相等的报文，称其为报文对,这两个报文到达同一点之间的时间间隔称为报文对间隔。考虑这个报文对在通路上的传输过程: 到达瓶颈信道之前，报文对间隔保持为 T；到达瓶颈信道时，信道的转发延时最大，$D_{b-1} = P / L_b > T$，这时将出现报文排队等候处理的现象；在此之后，不会再出现排队现象，报文对之间的时间间隔将保持为 D_{b-1},(如图 1 所示)，直至报文对到达宿主机。若宿主机在受到报文后立即给出响应，在测试主机观察到的报文对间隔仍为 D_{b-1}，而由 $D_{b-1} = P / L_b$,得 $L_b = P / D_{b-1}$，从而可求瓶颈信道的带宽。

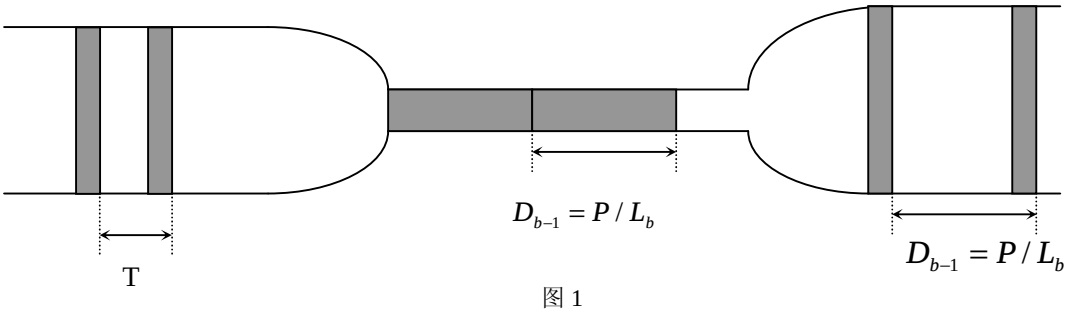


图 1

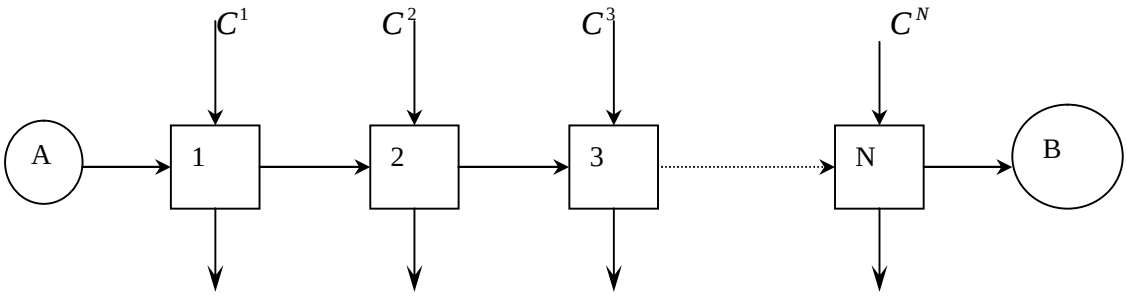
如果需要进一步发现网络通路的瓶颈位于何处，可采用如下做法。首先测出该通路的瓶颈带宽，然后利用连续发送的 ICMP 报文发现源点至宿点的逐跳路由，这样从离源点最近的一跳开始，测试源点到路由上每一跳的通路瓶颈带宽，直至测试结果等于整个通路的瓶颈带宽，这时最后一跳即为瓶颈信道。事实上，可以采用对分查找算法进一步提高效率,考虑到若测出的至某跳的瓶颈带宽值大于通路的瓶颈带宽，则瓶颈信道必在该跳的下游位置，若测出的至某条的瓶颈带宽值等于通路的瓶颈带宽，则瓶颈信道在该跳的上游位置或就是该跳本身。

3. 网络报文对方法的数学模型分析

4.1 测试原理的模型分析

1. 模型假设

如图所示，A、B 之间的通路经过信道 1、2、.....N,



假定报文在信道上的延迟＝报文大小/信道速率；

2. **符号约定：**

在讨论之前，做如下符号约定：

d_j^i:第 j 个报文在信道 i 处的离去时间;

s_j^i:第 j 个报文在信道 i 处的延迟；

s^b： 瓶颈信道的延迟;

$v^i(t)$:在 t 时刻信道 i 处的等待时间;

c^i:在信道 i 处由于干扰流量报文而带来的额外延迟；

A^i:在信道 i 处报文对的两个报文的到达时间差；

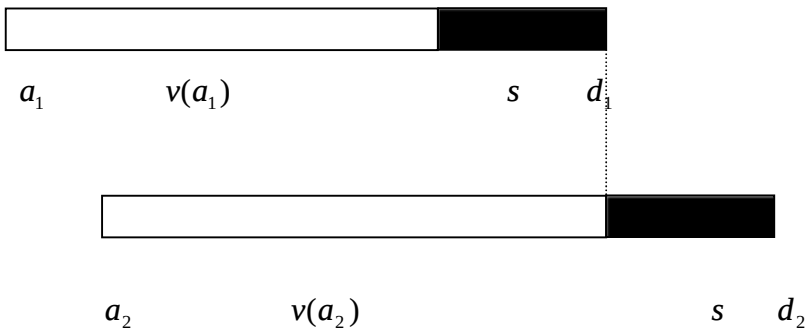
D^i:在信道 i 出报文对的两个报文的离去时间差

3. 模型的建立与求解

报文对测量方法的实质是通过观察到的两个报文的离去时间差 D_i 来获得 s_b,从而求出瓶颈信道的带宽，因而模型主要是关于 D_i 变量的讨论。

I 单一信道的情形

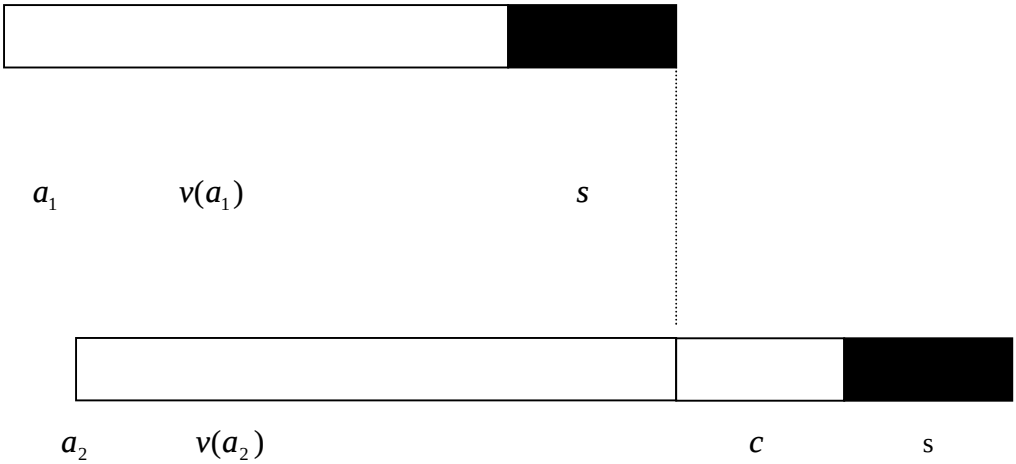
先从通路仅经过单一信道这种最简单的情形开始讨论。两个报文 P_1、 P_2 先后在 a_1、a_2 时刻到达信道 1，如下图



$$d_1 = a_1 + v(a_1) + s$$

$$d_2 = a_2 + v(a_2) + s$$

定义 $u = Max[v(a_1) + s - (a_2 - a_1), 0]$,注意到若无干扰流量的存在，有 $v(a_2) = u$,若而存在干扰流量时,$v(a_2) = u + c$



此时,

$$v(a_2) = u + c = Max[v(a_1) + s - (a_2 - a_1), 0] + c$$
$$= Max[d_1 - a_2, 0] + c$$

相应的

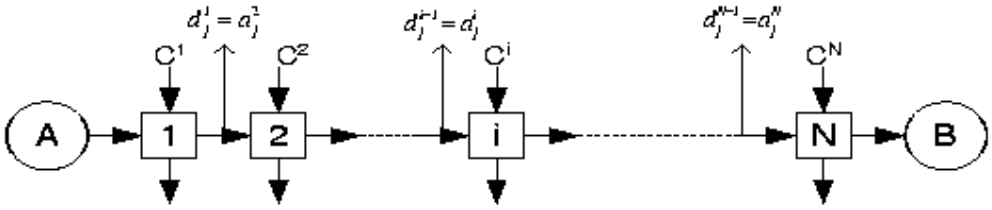
$$D = d_2 - d_1 = [a_2 + v(a_2) + s] - [a_1 + v(a_1) + s]$$
$$= [u + c] - v(a_1) + A$$
$$= Max[v(a_1) + s - A, 0] + c - v(a_1) + A$$
$$= Max[s, A - v(a_1)] + c$$

由上式，离去时间差 D 有两种可能的取值: $s + c$,或 $A - v(a_1) + c$，当无干扰流量即 $c = 0$ 时， D 的两种可能取值即简化为 s 或 $A - v(a_1)$

I N 条信道的情形

接下来讨论 N 条信道的普遍情形:

注意到此时某信道的到达时间差即为前一信道的离去时间差，即 $d_j^i = a_j^{i+1}$,(如下图所示)



定义 $f^i = c^i - v^i(a_1)$，同时令 $A^j = (a_2{}^j - a_1{}^j)$,根据上面的讨论，第一条信道处的报文离去时间差为

$$D^1 = Max(s^1, A^1 - v(a_1)] + c^1$$
$$= Max[s^1 + c^1, A^1 - v(a_1) + c^1]$$
$$= Max[s^1 + c^1, A^1 + f^1]$$

类似的，

$D^2 = Max[s^2 + c^2, A_2 + f^2]$ 而 $A_2 = D_1$,故

$$D^2 = Max[s^2 + c^2, Max[s^1 + c^1, A^1 + f^1] + f^2]$$
$$= Max[s^2 + c^2, s^1 + c^1 + f^2, A^1 + f^1 + f^2]$$

递推下去，可得，

$$D^N = Max[s^N + c^N, [s^{N-i} + c^{N-i} + \sum_{k=N-i+1}^{N} f^k]_{i=1}^{i=N-1}, A^1 + \sum_{k=1}^{N} f^k]$$

观察上式,N 条信道的离去时间差是 N+1 项的最大值，其中，前 N 项是干扰流量报文影响下的排队时间对最终离去时间的影响,最后一项代表着初始离去时间差(即主机发送两个报文的时间差)对最终离去时间的影响。

- **对结果的讨论**

当不存在干扰流量报文时，$c = 0$， $f^i = 0$,此时,

$D^N = Max[s^i |^N_{,i=1}, A^1]$,这时离去时间是所有信道延迟时间和初始离去时间的最大值，当初始离去时间小于瓶颈信道的服务时间时，所观察到的最终离去时间就是瓶颈信道的服务时间。这意味着若主机能足够快的在小于瓶颈信道服务时间内接连发送两个报文，则可以正确测量出瓶颈信道的带宽。

当存在干扰流量报文时，考察上式可以发现：若在瓶颈信道处未出现干扰报文，且 s^b 在 N+1 项取最大值时，则仍可获得正确结果，而不论非瓶颈信道是否存在干扰报文。事实上，只要在瓶颈信道上报文对发生毗连排队，且报文对在瓶颈信道的下游信道不再因拥塞而发生再次排队，而不论干扰报文的存在与否以及瓶颈信道上游信道是否发生过排队，测量结果总是正确的。注意到若是报文对在瓶颈信道的下游信道因拥塞而重新排队时，报文对的间隔将会被重置,这时的测量结果将错误的体现为出现重新排队信道的带宽值。

4. 报文对原理测试方法的实现时一些问题的解决方案

通过采用发送 ICMP ECHO 报文，然后接收 ICMP REPLY 报文的机制可以实现这种测量方法。但可能会遇到如下一些问题:

- 排队失败: 在瓶颈信道上报文对未发生排队等候处理现象；
- 干扰流量的存在:干扰流量是指测试时通路上的非测试报文的流量，这些流量报文可能正好出现于报文对的两个报文之间的位置，
- 测试报文丢包：测试报文在通路上由于某种特殊原因而丢失;
- 信道拥塞：当信道拥塞时，可能在非瓶颈信道上出现排队现象，从而改变报文对的时间间隔

以上这些问题的存在，都会影响测量结果，必须加以妥善解决。下面给出具体的解决方案：

- 排队失败的解决：在瓶颈信道上未出现排队的问题可通过两种途径来解决：1.缩短测试主机发送报文对的时间间隔，2.在条件允许的情况下尽可能发送尺寸较大的报文，大报文处理延迟较长，发生排队的可能性也相应较大
- 干扰流量存在的解决：干扰流量的报文若位置介于报文对之间，会带来额外的处理延时，增大报文对时间间隔，解决方法是接连发送大量的报文，构成多组不同尺寸的报文对，从而可以提高报文对之间无干扰流量报文的概率，最后在多组测量值由过滤过程进行过滤，选出合理测量结果。过滤过程的详细讨论将在后文中给出。实现时，报文尺寸按照 150%的比例增大，这样可以保证任何两个报文之间的大小不是整数倍关系。如果简单的将报文增大一倍，会存在如下问题：尺寸为 x 的报文对中有一个尺寸为 y 的干扰报文的测量结果，与尺寸为 2x 的报文对中有恰有两个尺寸为 y 的干扰报文的测量结果相同，从而使得过滤过程无法滤去这些受干扰的测量值。
- 测量报文的丢包的解决：网络传输中大尺寸的报文丢包可能性更大些，采用发送大量不同尺寸的报文可使该问题得以解决
- 信道拥塞的解决：当信道出现拥塞时，会导致在非瓶颈信道上的排队等候现象，这时测量值反映的将是出现拥塞的信道的速率。这一问题的解决交由过滤过程来处理。
- 过滤过程：报文对测试方法中最关键的一步，就是从大量的测试结果中过滤掉不合理的的结果，如那些排队失败的报文对，被干扰流量报文介入的报文对，以及在拥塞信道上重新造成排队的报文对给出的测量结果。图 3 中给出了针对一瓶颈信道为 56K 的 Modem 拨号信道的大量测量结果：

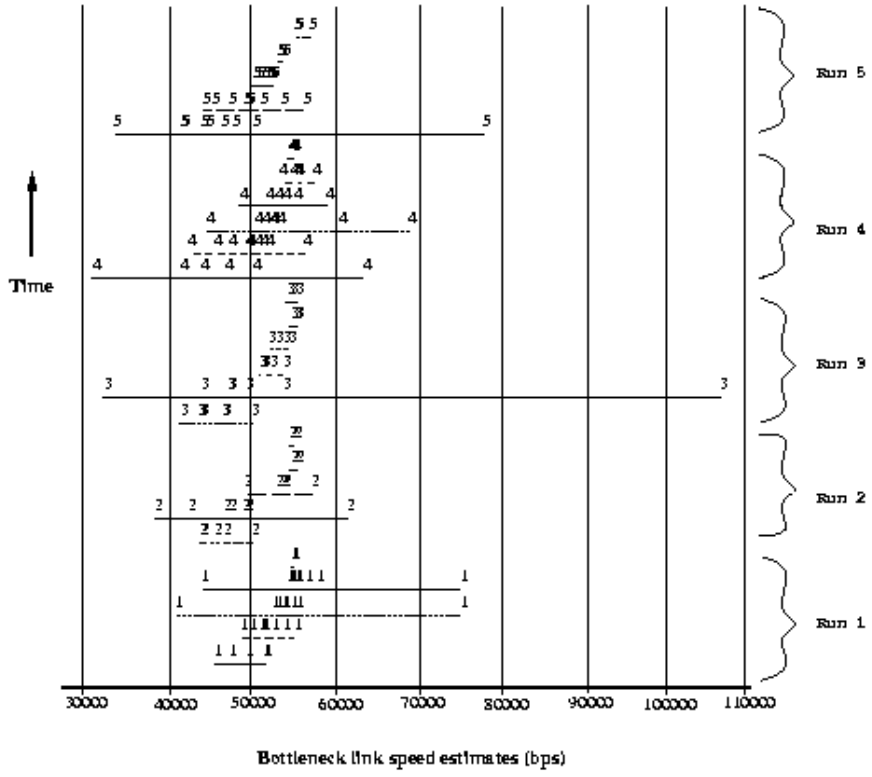


图 2

图中的数据是 5 轮(Run1 Run2… …Run 5)测试的结果，每轮的测试结果在图中用相应的数字(1、2、3、4、5)标出，每轮测量分别用 5 种尺寸的报文测量，在图中用直线标出，每种尺寸的报文各重复 9 次，即每条直线上会出现 9 个不同的值。图中，测试所用的报文尺寸按照由底部自顶部的方向逐渐增大。过滤过程基于如下事实，1)合理的测量值之间相关性好，在图中表现为聚集与某点附近，2)不合理的测量值之间缺乏相关性，例如被报文尺寸为 x 的干扰报文介入的报文对与被报文尺寸为 y (y≠x)介入的报文对均将导致测量值偏低，但测量值之间存在明显的差异。而被报文尺寸为 2x 的干扰报文介入的报文对与被报文尺寸为 y 的干扰报文介入的报文对之间的测量值差异将更大。根据此事实，直观上，测量结果应该选为所有测量值的最密集处，用数学语言来描述就是：测量值为b_1、b_2.....b_n ， 测量结果b 为使得$\sum_{i=1}^{n}(b_i-b)^2$ 最小的值。考虑到偏差越大的值的合理性越小，更为合理的做法是对$(b_i-b)^2$ 按照$(b_i-b)^{-1}$ 进行加权求和。

5. 实际测试结果

利用用报文对原理,在 CERNET 华东(北)地区网的实际环境中对各种典型的网络环境进行了测试。其结果见下表所示：

瓶颈信道类型	信道名称	跳数(Hops)	瓶颈信道速率	测量相对误差百分比的概率		
				<5%	<10%	<20%
DDN	网络中心——南京理工大学	5	64K	84%	93%	97%
以太网	网络中心内部局域网	3	10M	31%	73%	90%
卫星信道	清华大学——网络中心	4	2M	12%	20%	37%

由上表可见，若瓶颈信道为 DDN 时，测量值与真实值吻合的较好；对于以太网来说，由于带宽较大，受测试主机的时间精度的影响，相对误差较大一些；而对于卫星这种高吞吐量、大延时的特殊信道来说，测试结果存在严重的不合理性，宜改用其他方法。

6. 报文对原理测试方法的误差分析

- 主机的时钟精度的限制，这一点在测量高速瓶颈信道时显得尤为突出
- 测试中干扰流量报文的存在将可能影响测量结果，如前所述，干扰流量报文介于报文对之间从而带来的额外延时将使测量结果小于真实值
- 若被测通路上发生拥塞，可能导致报文对在非瓶颈信道上再次发生排队现象，这时测量结果反映的将是该信道的速率，通常使得测量结果明显大于真实值

7. 结论

基于报文对的网络瓶颈带宽测试方法是一种较为理想的测试技术，可用于测量通路的瓶颈带宽并发现瓶颈信道所在位置。它具有对如下一些特点：

- 工作在应用层，可作为用户的应用程序使用
- 对网络的冲击小，不应响网络的正常运作
- 测量时不需要安装特殊的服务器进行合作
- 适用于多种网络环境(速率几十 K 的广域网到十兆的以太网)，测量结果准确

但对于高速信道和特殊物理性质的信道(如卫星信道)，测试结果可能与真实结果具有较大的偏差。该方法产生的结果还可用于分布式信息服务的动态服务选择(Dynamic Service Selection)，帮助用户从一组服务器中选出服务质量最佳的服务器。

参考文献

[1]. Shikha Bahl Ashok Agrawala Analysis of Packet-Pair Scheme Estimating Bottleneck Bandwidth in a Netwok,1998

[2]. V.paxson Measure and analysis of End-to-End Internet Dynamics Ph.D. dissertation University of California and Berkeley, 1997

[3]. Robert L. Carter and Mark E. Crovella. Measuring Bottleneck Link Speed in Packet-Switched Networks. Technical Report BU-CS-96-006, Boston University, 1996.

[4]. Jean-Chrysostome Bolot. End-to-End Packet Delay and Loss Behavior in the Internet.In Proceedings of SIGCOMM 1993, pages 289-298. ACM SIGCOMM, August 1993.

[5]. Jean-Chrysostome Bolot. Characterizing End-to-End packet delay and loss in the Internet. Journal of High Speed Networks, pages:305-323, 1993.

[6]. Van Jacobson. Congestion Avoidance and Control. In Proceedings SIGCOMM '88 Symposium on Communications Architecturesand Protocols, pages 314-329, Stanford, CA, August 1988.