IPv6可用带宽测量方法的设计与实现

李智涛[1]，徐雅静[1]，刘利宏[2]，徐惠民[1]

(1. 北京邮电大学电信工程学院 北京 海淀区 100876; 2. 总装备部工程设计研究总院总体室 北京 朝阳区 100028)

【摘要】提出一种与现有带宽测量方法(包对法)不同的探测理论—— 基于链路状态的带宽测量方法。源端向目的端发送一系列小的探测报文，通过对探测报文的时延参数进行数理统计得到链路忙闲状态，进而获得链路可用带宽。采用此方法也可用于计算各种链路容量和链路空闲率。在实际测量中为解决探测报文与背景业务之间的相互影响，利用IPv6基本报头中流标签和业务类型字段设计专门的测试流和测试级。仿真证明方法正确。

关 键 词 可用带宽; 扩展报文头; IPv6; 链路状态; 时间戳

中图分类号 TP393 **文献标识码** A

Design and Implementation of Available Bandwidth Measurement in IPv6 Networks

LI Zhi-tao[1], XU Ya-jing[1], LIU Li-hong[2], and XU Hui-min[1]

(1. School of Telecommunication Engineering, Beijing University of Posts and Telecommunication Haidian Beijing 100876;
2. Beijing Special Engineering, Design Institute Chaoyang Beijing 100028)

Abstract A novel approach to available bandwidth measurement based on the link status is proposed, which is different from the packet train method. The source sends small packets as probes instead of sending packet pair/train back-to-back to achieve the link status and the available bandwidth. This method can measure the capacity and idle ratio of the targeted link. Meanwhile, in order to resolve the influence of cross traffic, the “test stream” and “test level” are designed using the corresponding fields in the IPv6 basic header. This approach is verified through simulations.

Key words available bandwidth; extension header; IPv6; link status; time stamp

随着Internet应用日益广泛，骨干链路和接入链路带宽成倍增长，网络测量开始受到关注。网络测量主要分为被动测量与主动测量。

近年来人们设计了大量的带宽测试算法和测试系统[1-7]。测量算法广义上主要有两类：(1) 数据包对(packet pair)算法，衍生算法有bprobe改进算法，TOPP算法、(改进型)Potential Bandwidth Filtering算法、Packet Tailgating算法、PBM算法等；(2) Pathchar算法，衍生的有非对称链路算法(适用于非对称链接网络的带宽测量)。

本文提出的测量方法通过探测源端随机向网络发送小探测报文，考察报文在网络中的延时特性，利用链路状态统计算法得到其忙闲状态，最后得到链路可用带宽，同时还可以分析各路由节点的流量变化。该方法较传统的包对法，不受源节点最大发送速率限制，对数据后处理和滤波要求较低。

1 基本概念

网络端到端路径是由从一台主机(源)到另一台主机(目的)传输数据包的一连串链路组成，记为P。端到端路径的可用带宽是指路径P所有链路中尚未被使用的最小链路容量，记为A。将决定最小可用带宽的一段链路称为紧链路(tight link)[8]。

设路径P跳数为H；C_i为链路i的最大传输量，可以表示为：

$$C=\min C_i \quad i=1,2,\cdots,H \tag{1}$$

如果U_i是链路i在一段时间间隔内的利用率($0\leqslant U_i\leqslant 1$)，则端到端路径可用带宽表示为：

$$A=\min[C_i(1-U_i)] \quad i=1,2,\cdots,H \tag{2}$$

本文采用文献[9]中的思想利用链路空闲率修正可用带宽的定义。

首先，给出链路可用带宽中“可用”的定义。

收稿日期：2007－04－27；修回日期：2008－01－12
基金项目：国家863计划(2006AA01Z235)；国家自然科学基金项目(90604019)
作者简介：李智涛(1978－)，女，博士生，主要从事IP网络流量、行为测量及QoS管理方面的研究.

定义1　某一链路“可用”就是该链路的开始节点处于空闲状态。可采用状态阶跃函数来描述节点在某一时刻所处的状态。

定义2　定义t时刻Node$_{i-1}$($1\leqslant i\leqslant H$)的状态函数如下：

$$\mathrm{Status}_{\mathrm{node}}^{i-1}(t)=\begin{cases}1 & \text{节点空闲}\\ 0 & \text{节点忙}\end{cases} \tag{3}$$

相应地，Node$_{i-1}$和Node$_i$之间的链路Link$_i$($1\leqslant i\leqslant H$)在某一时段$[t_1,t_2]$的链路空闲率表示为：

$$\mathrm{Free}_{\mathrm{link}}^{i}(t_1,t_2)=\frac{1}{t_2-t_1}\int_{t_1}^{t_2}\mathrm{Status}_{\mathrm{node}}^{i-1}(t)\mathrm{d}t \tag{4}$$

而Link_i在某一时段$[t_1,t_2]$的可用带宽$B_{\mathrm{aval}}^{i}(t_1,t_2)$可表示为：

$$B_{\mathrm{aval}}^{i}(t_1,t_2)=C_i\times\mathrm{Free}_{\mathrm{link}}^{i}(t_1,t_2) \tag{5}$$

式中　C_i为该链路的容量。

显然，当Link$_i$处于空闲状态时，对于一个新到来的报文，Node$_{i-1}$可以立即转发该报文。也就是说，当Link$_i$处于空闲状态时，报文在Node$_{i-1}$所经历的排队延时为0。考虑到实际应用中不存在“同时发生”的网络事件，因此有如下推论：对于一个新到来的报文，若该报文在Node$_{i-1}$所经历的排队时延为0，则说明在该报文被Node$_{i-1}$接收的时刻，Link$_i$处于空闲状态。

假设一条网络路径由N段首尾相连的存储转发链路组成(从Node$_0$到Node$_i$)，可用带宽最少的链路出现在网络链路的第b段，Link$_b$的容量为C_b，则该网络路径在某一时段$[t_1,t_2]$的可用带宽为：

$$\mathrm{Avail\_}bw_{0\sim N}=C_b\times\mathrm{Link\_Free}_b(t_1,t_2) \tag{6}$$

2　IPv6带宽测量方法

本文提出的测量方法正是基于上述理论，通过链路忙闲状态，获得链路可用带宽。那么，如何获得空闲状态呢？探测源端主动向网络发送带有IPv6时间戳扩展报文头的探测报文，逐跳记录路由器的当前时间。多个探测结果通过改进的链路状态统计算法处理，消除处理时延、传播时延以及路由器时钟扭曲带来的偏差，得到探测分组单跳排队时延，进而转化为节点忙闲状态，获得链路的可用带宽。同时为了消除探测报文对背景流量的影响，利用IPv6报头中流标签和业务流字段，设计专门的测试级，保证探测报文序列路径一致并且不被其他竞争流分离。该方法要求被测路径中路由器支持IPv6时间戳扩展报文即可，对目的端设备无特殊要求。

下面首先介绍测试报文的设计，接着分析改进型链路状态统计算法，最后介绍测试级的设计。

2.1　测试报文设计

(1) 带有IPv6时间戳扩展报文头的ICMP报文。采用带有IPv6时间戳扩展报头的ICMPv6 Echo Request(请求回显)探测报文，目的端可以是任意支持ICMPv6的节点。目的端收到此探测报文后，按要求生成ICMPv6 Echo(回显)报文，并将原报文中的IP扩展报头数据复制到回显报文中，发回给源端。

(2) 带有IPv6时间戳扩展报头的TCP/UDP探测报文，这种探测报文在网络上的行为与其他TCP/UDP6行为类似，避免了路由器对ICMPv6报文的特殊处理，可以更加准确地反映网络的动态特性。源端向目的端发送带有IPv6时间戳扩展报头的TCP/UDP报文后，目的端接收这些报文并提取时间戳信息做进一步处理。下面介绍IPv6时间戳扩展报头的设计。

由于IPv6协议中没有IPv4协议的“时间戳选项”，因此设计IPv6时间戳扩展报文头，同时要求路由器支持对时间戳扩展头的处理。

IPv6扩展头通过IPv6头中“下一报头”字段来进行标识。根据国际组织IANA所给出的“Protocol Number”列表，目前未指派的编号是138-252[10]。对于时间戳扩展头，使用138来标识。根据RFC2460对IPv6扩展报头格式规定，定义IPv6逐跳时间戳选项扩展头格式如图1所示。

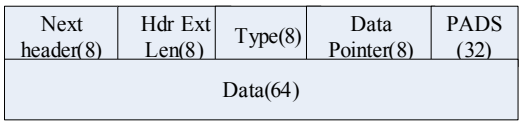

图1　IPv6时间戳选项报文格式

图1中：

(1) 下一报头(next header)　占用1字节，所有IPv6扩展头均包含此字段，用于标识下一头类型。

(2) 扩展头长度(header extension length)　占用1字节，用以标识时间戳选项扩展头长度，该长度以8字节为单位，不包含扩展头第一个8字节，即如果扩展头只有8字节长，该字段为0。此字段限制了扩展头最多为2 048字节。

(3) 类型(type)　占用1字节，标识时间戳选项头的类型，根据RFC2460建议，Type字段的高两位值为00，表示“若IPv6节点不支持该选项类型，则跳过本扩展头处理下一个扩展头”，本文以4表示“路由器时间戳”，Type字段格式为00000100。

(4) 数据指针(data pointer)　占用1字节，标识当前已经记录的时间戳数量，同时表明下一个时间戳存放在扩展头中的位置。Data Pointer初始值是0，当带有时间戳扩展头的IPv6报文经过路由器时，路由

器在“Data”字段写入时间戳，同时将Data Pointer值加1，除非“Data Point值已经是255”(即超过可记载最大时间戳数量)。路由器记录时间戳位置(该位置的起始地址是从时间戳扩展报文头的首部算起第(8×Data Pointer+1) 字节)的计算方法如下：

(5) 填充项(PADS) 占用4字节，仅用于填充，保证时间戳报文头在去除“Data”字段后长度是8字节的整数倍。

(6) 数据(data) 扩展头中数据，记录报文通过路由器的具体时间(时间戳)。时间戳是从UTC午夜开始到当前时间所经过的微秒数每个时间戳占用8字节，受“Hdr Ext Len”限制，最多记255个时间戳。

2.2 改进型链路状态统计算法及补偿算法

探测源端主动向网络发送如2.1中所述的探测报文，获得报文的单跳排队时延，利用改进型链路状态统计算法获得链路状态。通过大量仿真发现，对于吞吐量大，链路利用率变化显著的情况，如果只是简单地按照式(3)来统计链路忙闲状态不够准确。本文采用多阶量化算法，将链路状态从(0, 1)这个二阶量改成多阶量，得到的链路利用率更加贴近真实值。

记第i个探测报文的排队时延为$T_q^{(i)}$，设N个连续探测报文中最大的排队时延$T_{q\max}=\max(T_q^{(i)})i=1,2,\cdots,N$。根据探测精度的要求，将$T_{q\max}$进行$k$阶均匀量化得到序列$\{t\}=t_0,t_1,\cdots,t_j,t_{j+1},\cdots,t_{k-1}$。

定义k阶序列$\{s\}=s_0,s_1,\cdots,s_j,s_{j+1},\cdots,s_{k-1}$表示链路状态，其中$s_i=i/(k-1)$。$t_j$与$s_j$的映射关系可以通过线性或非线性函数定义，也可通过映射关系表定义，本方法即采用区间映射的方法获得第i个探测报文的链路状态S_i，即：

$$S=\{s\mid T\in[t_m,t),m=0,1,\cdots,k-2\}\cup\{s\mid T=t\} \quad (7)$$

$S_i(i=1,2,\cdots,n)$表示每次采样获得的链路忙闲状态，最后对忙闲状态通过统计窗口来处理，得到可用带宽。即在W时间内，以频率f对链路的忙闲状态持续采样，样本数为$n=Wf$；链路空闲率为：

$$\mathrm{Free}_{\mathrm{link}}^{i}=n^{-1}\sum_{i=1}^{n}S_i \quad (8)$$

在仿真中发现，网络发生拥塞时，报文会丢失，随着丢包率的提高，探测报文丢失数量也随之增加，无法获得时间戳。本文提出链路状态补偿算法预测丢包情况下链路的忙闲情况。

从网络流量的自相似特性可知，在短时间内网络状态具有较强的相关性，故可以依据某一时刻前后相邻时刻的状态来预测该时刻网络的状态，即：

$$S(t)=A\times F(S(t-\Delta t),S(t+\Delta t))$$

式中 $S(t)$、$S(t-\Delta t)$、$S(t+\Delta t)$分别为t、$t-\Delta t$、$t+\Delta t$时刻网络状态。

若探测包间隔足够小，可以根据探测包P_i前后两个包P_{i-1}、P_{i+1}来判断丢失探测包P_i所处单位时间内的链路状态。由于链路忙闲状态是单跳排队时延$T^{(i)}$[11]的量化值，故直接取其作为参数，得到：

$$T^{(i)}=\alpha T^{(i-1)}+(1-\alpha)T^{(i+1)} \quad (9)$$

式中 α为补偿系数；$T^{(i-1)}$和$T^{(i+1)}$分别表示报文P_{i+1}和P_{i-1}探测的单挑时延。

通过预测的单跳时延就可以进一步计算得到信道利用率。

2.3 测试级的设计

对于主动测量中普遍存在的测试报文路径不一致以及与背景业务之间相互影响的问题，本文通过IPv6报文头中流标签、业务类型字段定义专门的测试流和测试级来解决。

IPv6包头定义了20 bits的流标签字段和8 bits的流量类型字段。流标签字段用于源节点标识IPv6路由器需要特殊处理的包序列。本方法中将所有探测报文的流标签字段定义为统一值，报文在转发过程中选择固定且唯一的路径。同时为了解决探测报文与背景流相互影响的问题，利用IPv6报头中流量类型字段，增加特殊的测试流级别(简称测试级)。测试级被定义为IPv6报文中的最低优先级，源节点和转发路由器只有在没有其他数据进行处理的时候才会处理测试报文。所以测试报文不会对已有业务产生影响。因此，测试时可以用很小间隔发送大量探测报文得到测试结果，提高测试效率。

3 仿真及分析

3.1 仿真

仿真配置环境如图2所示，X、Y、Z的链路容量分别是80、50、80 Mb/s；虚线代表背景流量，分别流经链路X、Y和Z，包括one-hop persistent，path persistent。首先考察在链路利用率为20%、80%和50%时包对法的测量结果。

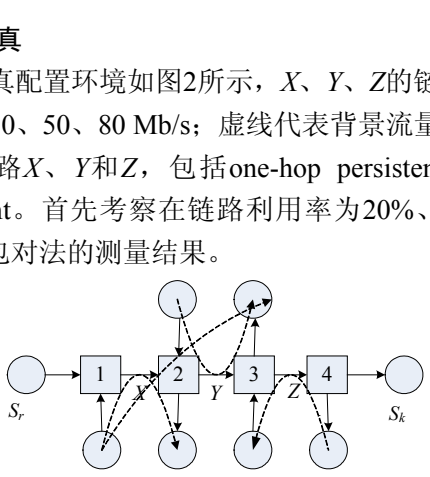

图2 仿真配置图

对于本文提出的方法，从源端向目的端发送带有IPv6时间戳选项的ICMPv6 Echo Request报文，目

的端接收此报文后，生成ICMP Echo(请求回显)报文，并将原报文中的IPv6扩展报文头中数据复制到回显报文中，发回源端。发送间隔为10 ms，背景流量为pareto分布(α =1.9)，考察在不同可用带宽下的测量结果。

3.2 仿真结果分析

图3为包对列法(N=3)在3种链路利用率下1 500次试验中的结果分布情况。可以看出包对列法测量结果逐渐收敛，集中在文献[8]中提到的ADR (asymptotic dispersion rate)附近。但是，无论链路负载情况如何，包对列法无法准确测出可用带宽。

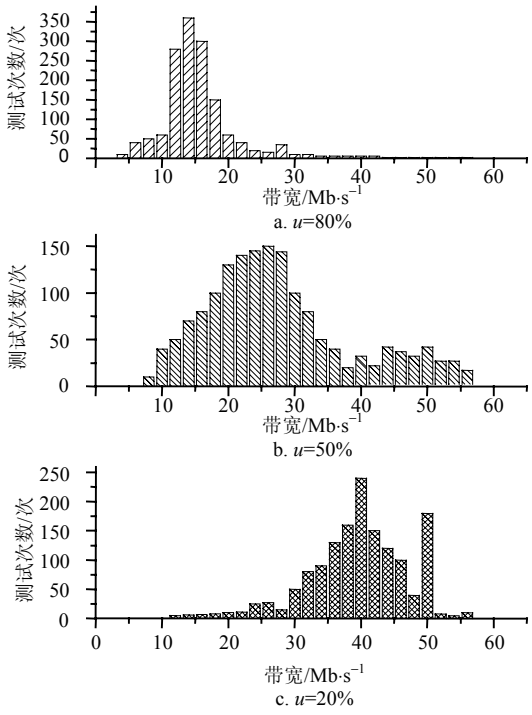

图3　包对法仿真结果

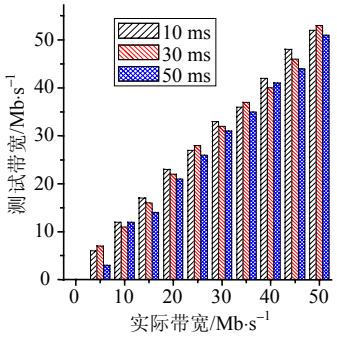

图4　状态法仿真结果

图4给出了实际可用带宽值在0、5、10、15、20、25、30、35、40、45、50 Mb/s时，滑动窗口分别为10、30、50 ms的测量结果。可以看出采用不同的滑动窗口均能够反映出链路可用带宽的真实值。其中探测结果偏差产生的原因主要在于：(1) 处于探测分组间隙处的链路情况无法探测到；(2) 链路量化阶数不够细致，导致状态统计误差；(3) 链路利用率很高时，报文丢失会导致探测结果不准确。

4 结　论

本文提出一种适用于IPv6网络的可用带宽测量方法。此方法不同于传统的包对法，通过发送小的探测报文获得链路的忙闲状态，通过统计获得链路的可用带宽。这种方法不受测试报文发送速率的限制，同时发送报文数小，对网络中已有业务影响小。仿真验证测量方法可行。

下一步将研究把该方法应用于无线环境的带宽测量。

参　考　文　献

[1] LIU Min, SHI Jing-lin, LI Zhong-cheng, et al. A new end-to-end measurement method for estimating available bandwidth[C]//The 8th IEEE International Symposium on Computers and Communications. San Francisco: IEEE Press, 2003: 1393-1400.

[2] LAI K, BAKER M. Nettimer: a tool for measuring bottleneck link bandwidth[C]//The Proceedings of USITS ’01. San Francisco: IEEE Press, 2001.

[3] LAI K, BAKER M. Measuring link bandwidths using a deterministic model of packet delay[C]// SIGCOMM 2000. Stockholm: ACM Press, 2000.

[4] MAH B A. Pchar: a tool for measuring internet path characteristics [EB/OL]. [2007-04-10]. http://www. employees. org/ bmah/ software/ pchar/.

[5] CARTER R L, CROVELLA M E. Measuring bottleneck link speed in packet switched networks[J]. Performance Evaluation, 1996, (27-28): 297-318.

[6] SAROIU S , KRISHNA G P, STEVEN D. Gribble SProbe: a fast technique for measuring bottleneck[J/OL]. [2007-03-24]. http://sprobe ce. Washing ton. Edu/sprobs. Ps.

[7] BESTAVROS J BYERS, HARFOUSH K. Inference and labeling of metric-induced network topologies[C]//IEEE INFOCOM 2002. New York, USA: IEEE Press, 2002: 628-637.

[8] DOVROLIS C, RAMANATHAN P, MOORE D. What do packet dispersion techniques measure?[C]//IEEE INFOCOM 2001. Anchorage, USA: IEEE Press, 2001: 905- 914.

[9] LIU Min, LI Zhong-cheng, GUO Xiao-bing, et al. An end-to-end available bandwidth estimation methodology[J]. Journal of Software, 2006,17(1): 108-116.

[10] RFC 2460. Internet protocol version 6 (IPv6) specification[s].

[11] 崔毅东, 林　宇, 徐雅静, 等. 新的基于逐跳时间标签的链路利用率测量方法[J]. 北京邮电大学学报, 2006, 29(2): 5-9.
CUI Yi-dong, LIN Yu, XU Ya-jing, et al. An approach to link utilization measurement based on per-hop time stamp[J]. Journal of Beijing University of Posts and Telecommunications, 2006,29(2): 5-9.

编　辑　熊思亮