

一种测量任意链路可用带宽的方法^{*}

何莉⁺, 余顺争

(中山大学 电子与通信工程系, 广东 广州 510275)

Methodology for Measuring Available Bandwidth on Arbitrary Links

HE Li⁺, YU Shun-Zheng

(Department of Electronics and Communication Engineering, SUN YAT-SEN University, Guangzhou 510275, China)

+ Corresponding author: E-mail: heli_hefen@hotmail.com

He L, Yu SZ. Methodology for measuring available bandwidth on arbitrary links. Journal of Software, 2009, 20(4):997-1013. <http://www.jos.org.cn/1000-9825/3306.htm>

Abstract: The available bandwidth measurement is important for many Internet applications, such as network behavior analysis, quality of service (QoS) verification, and so on. Existing available bandwidth measurement tools mainly measure the available bandwidth for a whole network path, and provide information about the tight link(s) other than vital links. Therefore, this paper presents a novel algorithm called LinkPPQ (trains of pairs of packet-quartets used to measure available bandwidth of arbitrary links), which uses trains of pairs of packet-quartets to measure the available bandwidth of any link along the path, and track the variety of the cross-traffic on it. This paper studies the performance of LinkPPQ in both simulation circumstances and the laboratorial network. Simulation results show that LinkPPQ can accurately measure the available bandwidth of each link on the paths that have one narrow link or multiple narrow links, under different cross-traffic loads. Most measurement errors are under 30%, and the results are stable. The laboratorial experimental results show that LinkPPQ can accurately measure the available bandwidth in the following situations: a) measuring a link with capacity 100Mbps from a link with capacity 10Mbps; b) monitoring the link with capacity that is ten times of the narrow link next to it on the path; c) estimating the available bandwidth of the narrow links on a path having multiple narrow links.

Key words: link available bandwidth; active probing; network monitoring; Internet measurement; performance evaluation

摘要: 可用带宽测量对于网络行为分析、网络服务质量(QoS)的验证等有很重要的作用. 现有可用带宽测量工作主要集中在端到端路径可用带宽测量,仅提供路径上承压链路(tight link)的信息,而不能提供其他关键链路的信息.为此,提出一种新颖的链路可用带宽测量算法 LinkPPQ(trains of pairs of packet-quartets used to measure available bandwidth of arbitrary links),它采用由四探测分组结构对构成的探测序列,能够测量网络中任意链路的可用带宽,并跟踪该链路上背景流的变化.在仿真环境和实际网络环境下研究了 LinkPPQ 的性能.仿真结果表

* Supported by the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z143 (国家高技术研究发展计划(863)); the Key Program of NSFC-Guangdong Joint Funds of China under Grant No.U0735002 (国家自然科学基金-广东联合基金重点项目)

Received 2007-12-15; Accepted 2008-02-27

明,在几种不同背景流场景下,对于具有单狭窄链路的路径和具有多狭窄链路的路径,LinkPPQ 都能够对各个链路的可用带宽进行有效的测量.绝大多数情况下测量误差小于 30%,且具有较好的测量平稳性.实验网的实验结果也表明,LinkPPQ 可以准确测量以下几种情况下的链路的可用带宽:a) 从容量为 10Mbps 的链路准确地测量一条 100Mbps 链路的可用带宽;b) 准确测量容量 10 倍于紧邻其后狭窄链路的容量的链路的可用带宽;c) 准确测量具有多狭窄链路的路径上各狭窄链路的可用带宽.

关键词: 链路可用带宽;主动探测;网络监控;因特网测量;性能估计

中图法分类号: TP393 文献标识码: A

可用带宽测量对许多网络应用和协议有很重要的作用.流媒体的速率控制、端到端的接入控制、服务器的动态选择、覆盖网络(overlay network)的路由选择、拥塞控制以及服务质量 QoS 的验证等都需要网络可用带宽测量的支持.虽然网络拓扑的改变和网络流量的瞬间变化使得在互联网中进行可用带宽测量工作变得困难,但是网络背景流在至少数分钟内是平稳的^[1],以及网络中绝大多数端到端路径采用单条路由并且持续数天或数星期^[2],这使得可用带宽测量成为可能.

网络测量分为主动测量和被动测量.被动测量一般在网络中设置监测点,捕捉流经监测点的网络背景流,或者查询网络设备的管理信息库(management information base,简称 MIB)获得网络各接口的状态信息^[3].网络行为分析系统通常应用于企业内部互联网中,使用 cFlow^[4],sFlow^[5]或 Cisco 的 Netflow^[6]采集支持对网络流的统计或配置了简单网络管理协议(simple network management protocol,简称 SNMP)的路由器上的流量数据,进而采用特征检测和异常检测机制从网络流量数据中识别感兴趣的网络行为.被动测量由于能够提供流的变化和分布信息,特别适合于流量工程和流的建模.然而,在高速发展的互联网上存储流量数据需要巨大的空间,并且由于企业竞争和保密的需要,网络服务提供商(Internet server provider,简称 ISP)不会向公众提供查询路由器上流量信息的权限,也不允许在网络内部随意设置监测点.因此,在互联网中要实施网络行为分析,需要采取主动测量的方式.主动测量一般在网络的边缘引入探测分组.由于网络上背景流使得探测分组延迟、丢失或乱序,端主机在网络边缘分析穿过网络的探测分组携带的信息获得网络的状况.主动探测的优点是不需要网络中间路由器提供特殊的权限,在网络的边缘就获取中间网络的信息,尤其适合于观察网络中端到端应用的 QoS.主动测量的缺点是引入的探测分组流可能会影响网络中的背景流,可能增加网络的负荷.因此,设计主动测量算法时需要考虑探测流的影响.

可用带宽测量一般属于主动测量.现有的研究成果主要集中在端到端路径的可用带宽测量方面,而针对链路的测量工作主要集中在容量测量上.端到端的可用带宽测量^[7-14]给出整条路径的信息,实际上提供的是承压链路(可用带宽最小的链路)的信息,不能提供非承压链路的信息.然而,获取链路的却不仅仅是路径的可用带宽在许多网络应用中有重要作用.在互联网中进行网络行为分析,需要采取主动的方式获得每条链路的流量.对于服务器负载均衡的实现,一个值得考虑的因素是服务器周围网络具有较大的可用带宽.在 P2P 应用中,端主机为快速获取数据,需要选择周围网络负载比较低的节点提供资源.由于任意两条路径可能交叉在一起并且拥有相同的承压链路,由多条路径构成的网络的总可用带宽不能简单地通过测量每条路径的可用带宽获得,因此,需要对服务器周围网络上的每条链路进行测量,或者通过测量某些关键的链路的可用带宽来估计总可用带宽.

这些因素促使我们做了如下几方面的工作:

第一,我们提出了一种新颖的链路可用带宽测量算法 LinkPPQ(trains of pairs of packet-quartets used to measure available bandwidth of arbitrary links),它采用特殊设计的探针,能够定位并测量路径上任意链路的可用带宽.这显著不同于现有的端到端路径可用带宽测量算法.我们从理论上给出了探测分组长度选择规则,并分析了 LinkPPQ 的有效性.

第二,我们在 NS-2^[15]仿真环境下观察影响 LinkPPQ 测量性能的两个参数:探测分组大小和数量,并根据局部路径容量与被测链路容量的大小关系,分两种情况研究参数的取值.仿真结果被用于指导实际网络的测量.

第三,我们在 NS-2 仿真环境下选取两条典型的路径:一条是具有单狭窄链路的路径;另一条是具有多狭窄

链路的路径,通过引入实际环境中的背景流,改变路径上各链路的利用率,构造不同的流量场景,对路径上的每条链路的可用带宽进行测量.仿真结果表明,LinkPPQ 测量结果都比较准确,绝大多数情况下测量误差小于 30%,并具有良好的稳定性,即,绝大多数情况下测量平稳性大于 80%.

第四,我们在实验网上构建了 3 条典型的路径验证 LinkPPQ 的性能.LinkPPQ 的探测流经过一条容量为 10Mbps 的链路准确地测量了紧邻其后的一条容量为 100Mbps 的链路的可用带宽;准确地测量了容量为 10Mbps 的狭窄链路前面的容量为 100Mbps 的链路的可用带宽,而此时狭窄链路的背景流是变化的;还准确地测量了一条路径上多条狭窄链路的可用带宽.这 3 种情况都是现有路径可用带宽测量算法所无法测量的.

本文第 1 节介绍可用带宽的基本概念以及相关工作.第 2 节提出链路可用带宽测量算法 LinkPPQ,并进行理论分析.第 3 节描述仿真环境,研究探测分组长度和数量对 LinkPPQ 性能的影响.第 4 节首先在仿真环境下,构造两条典型的路径,验证 LinkPPQ 对路径上每条链路进行测量的性能;其次在实验网上构建 3 条典型路径,验证 LinkPPQ 在实际环境中的测量性能.第 5 节总结全文.

1 基本概念和相关工作

一般地,一条网络路径由一系列将分组从发送端存储转发到接收端的链路组成,可描述为 $P=(L_1, L_2, \dots, L_N)$. 路由器 R_{i-1} 与 R_i 之间的链路 L_i 的容量 C_i 为该链路上在 IP 层能够传输的最大吞吐量.路径的容量由容量最小的链路决定,可表示为 $C=\min(C_1, C_2, \dots, C_N)$.如果链路 L_i 在一段时间内的平均利用率为 $u_i(0 \leq u_i \leq 1)$,相应地,背景流平均强度为 $\lambda_i=u_i \times C_i$,链路 L_i 的可用带宽定义为 $A_i=C_i-\lambda_i$,而路径的可用带宽为 $A=\min(A_1, A_2, \dots, A_N)$.容量最小的链路称为狭窄链路(narrow link),可用带宽最小的链路称为承压链路(tight link)^[16].显然,狭窄链路不一定是承压链路,一条网络路径上可能存在多条狭窄链路或多条承压链路.局部路径 $P_{i,j}(j>i)$ 由链路 L_i, L_{i+1}, \dots, L_j 构成, $C_{i,j}=\min(C_i, C_{i+1}, \dots, C_j)$ 表示 $P_{i,j}$ 的容量, $B_{i,j}=\operatorname{argmin}_{i \leq k \leq j} \{C_k\}$ 代表 $P_{i,j}$ 的狭窄链路集,包含一条或多条狭窄链路, $A_{i,j}=\min(A_i, A_{i+1}, \dots, A_j)$ 表示 $P_{i,j}$ 的可用带宽,它不一定是 $B_{i,j}$ 的可用带宽,即不一定有 $\operatorname{argmin}_{i \leq k \leq j} \{A_k\} \subset B_{i,j}$.

目前,在端到端路径可用带宽测量方面有许多成果.Keshave^[17]于 1991 年设计了一种分组对速率探测算法以及一种基于速率的流控机制,但需要假设路由器采取公平调度机制,这不适合于通常的路由器采取 FIFO(first in first out)队列机制的网络.Cprobe^[18]和 pipechar^[19]的潜在假设是分组串的间隔反比例于可用带宽.然而,文献[20]指出,分组串的间隔测量的不是可用带宽,而是一种被称为 ADR(asymptotic dispersion rate)的度量,它低于路径的容量,但不是可用带宽.

根据探测方式,目前的可用带宽测量模型分为探测速率模型(probe rate model,简称 PRM)和探测间隔模型(probe gap model,简称 PGM).

PRM 是基于自诱导拥塞(self-induced congestion)的概念.一般地,如果源端发送探测流的速率小于路径可用带宽,那么探测流到达接收端的平均速率将等于发送端的发送速率;反之,探测流将会拥塞网络路径,网络反作用于探测分组,使其产生延迟,所以探测流到达接收端的平均速率将小于发送速率.因此,PRM 在测量开始阶段寻找探测流发送速率和到达速率开始匹配(相等)的转折点,然后将转折点对应的探测分组平均到达速率作为路径可用带宽的估计.PRM 不需要事先获得路径的容量.这类典型的算法有 TOPP^[13],Pathload^[12],YAZ^[14],abget^[21],Pathchirp^[10],PTR^[7].

PGM 分析两个连续的分组的发送间隔与经过路由器后到达接收端的间隔的关系,以估计可用带宽.PGM 通常假设路径上仅有一条瓶颈链路,同时是狭窄链路和承压链路,由于此链路对分组对的间隔变化影响最大,PGM 把此路径抽象成一条链路.两分组以一定的间隔由发送端先后传往接收端,如果在第 1 个分组离开瓶颈链路的缓冲区队列之后、第 2 个分组到达之前缓冲区队列非空,那么两分组到达接收端的间隔与发送间隔呈线性关系,能够体现瓶颈链路上背景流的信息,进一步获得路径的可用带宽.由于没有 PRM 所需要的迭代过程,PGM 的测量相对简单,但需要事先获取路径的容量.这类典型的算法有 IGI^[7],Delphi^[9],spruce^[11].

目前,国内在可用带宽测量方面也有一些有意义的工作.COPP^[22]属于 PRM 模型,区别于 TOPP^[13],COPP 发送分组对内间隔逐渐减小的分组对序列,在接收端从每组序列中找出所有成为转折点的分组对,然后根据分组

对所受干扰的不同程度给予不同的权重,进而获得路径的可用带宽.文献[23]基于蒙特卡洛(Monte Carlo)随机抽样的思想,通过确定承压链路的链路空闲率,获得路径的可用带宽.文献[24]改进了 IGI 和 PTR 算法,提高测量的准确度和平稳性,增强了跟随背景流变化的能力.

与可用带宽测量紧密相关,Pipechar^[19],Bfind^[25],pathneck^[26]以及 STAB^[27]等算法能够用于网络路径上承压链路的定位.

目前的可用带宽测量算法主要针对路径的测量,给出的是整条路径的总的信息,不能提供承压链路之外其他链路的信息.然而,能够定位并且测量其他关键链路的可用带宽有着重要的作用,这促使我们设计一种能够对任意链路进行可用带宽测量的算法.

2 链路可用带宽测量算法 LinkPPQ

本节首先描述 LinkPPQ 采用的探针分组结构,给出链路可用带宽估计的方法,然后从理论上分析 LinkPPQ 的有效性.

2.1 探 针

探测分组 p 有几个特征:分组大小,用 $s(p)$ 表示;分组的目的地链路,用 $D(p)$ 表示. $D(p)$ 一般通过结合分组的目的地和 TTL 域来设置,其中,目的地一般设置成接收端的地址.假如一条网络路由由 N 条链路组成. $D(p)=L_i$ ($i < N$) 表示分组 p 由发送端向接收端传输,在即将离开链路 L_i 时,分组 p 会因为 TTL 超时而被丢弃,此时,分组 p 的 TTL 域设置为 i ;如果希望分组到达接收端,则设置 TTL 域为超过 N 的值.分组 p 和 q 在链路 L_i 的到达(离开)间隔 Δ_m^i (Δ_{out}^i) 是指 p 和 q 的最后一个比特到达(离开)此链路对应的输出缓冲区队列的时间差.

假设路径 P 由 N 条链路组成,分别为 L_1, L_2, \dots, L_N ,链路的容量分别为 C_1, C_2, \dots, C_N .我们的目标是测量链路 L_k 的可用带宽 A_k ,其中, $2 \leq k \leq N-1$.我们设计了由两个四探测分组结构 $\{p_1, m_1, q_1, m_2\}$ 和 $\{p_2, m_3, q_2, m_4\}$ 构成的探针,如图 1 所示,其中,分组 m_1, m_2, m_3, m_4 称为标志分组(marker packet),而分组 p_1, q_1, p_2, q_2 称为定速分组(pacesetter packet).相邻两个四探测分组结构之间的间隔 T_{inter} 设置得足够大,以避免四探测分组结构相互影响.探针中, $\{p_1, m_1, q_1, m_2\}$ 用于测量两相邻探测分组进入 L_k 的间隔,而 $\{p_2, m_3, q_2, m_4\}$ 用于测量两相邻探测分组离开 L_k 的间隔.LinkPPQ 通过分析探测分组到达和离开 L_k 的间隔,计算 L_k 的可用带宽.探针中两个四探测分组结构之间的共同之处是:

- 定速分组 p_1 与 p_2 的长度相同,即 $s(p_1)=s(p_2)=s(p)$;
- 定速分组 q_1 与 q_2 的长度相同,即 $s(q_1)=s(q_2)=s(q)$,且 $s(q) \leq s(p)$;
- 标志分组 m_1, m_2, m_3, m_4 具有相同长度并且都到达接收端 P_d ,即 $s(m_i)=s(m)$, $D(m_i)=L_N$,其中, $i=1, 2, 3, 4$;
- 探针中,每个四探测分组结构中的 4 个分组以一个紧挨着一个(也称背靠背)的形式进入 L_1 ,如图 1 所示,于是有 $T_{intra}=[s(q)+s(m)]/C_1$.

不同之处是,定速分组 p_1 和 q_1 、 p_2 和 q_2 的目的地链路分别为 L_{k-1}, L_k ,即 $D(p_1)=D(q_1)=L_{k-1}$, $D(p_2)=D(q_2)=L_k$.通过设置探针中定速分组的目的地链路,能够定位路径上的任意链路.

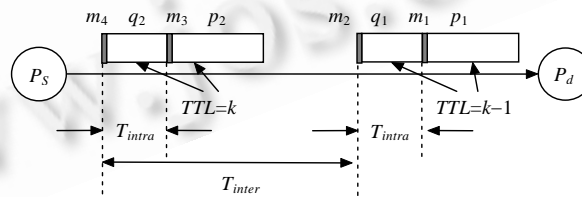


Fig.1 Structure of the probe consisting of two packet-quartets

图 1 由两个四探测分组结构构成的探针的结构

类似的分组结构被用于测量链路的容量^[28-30],然而,我们使用由两个四探测分组结构构成的探针来估计链路的可用带宽.

探针 $\{p_1, m_1, q_1, m_2\}$ $\{p_2, m_3, q_2, m_4\}$ 在路径上的传输过程如图 2 所示. 当探针离开 L_{k-1} 时, 分组 p_1 和 q_2 因 TTL 超时而被丢弃, 其余探测分组继续往 L_k 方向传输, 其中, m_1 和 m_2 以间隔为 $\Delta_{m,1}^k$ 到达 L_k , 而 m_3 和 m_4 到达 L_k 的间隔为 $\Delta_{m,2}^k$. 当探针离开 L_k 时, 分组 p_2 和 q_1 因 TTL 超时而被丢弃, 标志分组 m_1 与 m_2 、 m_3 与 m_4 分别以间隔 $\Delta_{m,1}^{k+1}$ 、 $\Delta_{m,2}^{k+1}$ 离开 L_k (到达 L_{k+1}), 并最终在 L_{k+1} 以间隔 $\Delta_{out,1}^N$ 、 $\Delta_{out,2}^N$ 到达接收端.

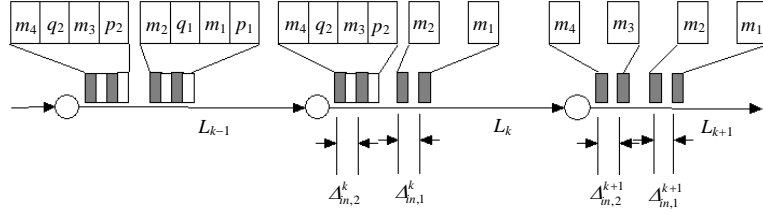


Fig.2 Transmission process of the probe consisting of two packet-quartets

图 2 由两个四探测分组结构构成的探针的传输过程

由于背景流是随机变化的, 需要向被测链路 L_k 发送由多个探针构成的探测序列, 以捕捉 L_k 上背景流的信息. 一个四探测分组结构可等效成一个探测分组对. m_3 和 m_4 到达和离开 L_k 的间隔 $\Delta_{m,2}^k$ 、 $\Delta_{m,2}^{k+1}$ 分别对应等效分组对在 L_k 的输入和输出间隔. 在一定的条件下, 等效分组对经过 L_k 时在同一个缓冲区队列中, $\Delta_{m,2}^{k+1}$ 与 $\Delta_{m,2}^k$ 呈线性关系^[9,11,13,31], 即

$$E[\Delta_{m,2}^{k+1}] = \frac{[s(q_2) + s(m_2)] + E[\Delta_{m,2}^k] \lambda_k}{C_k} \quad (1)$$

其中, λ_k 是链路 L_k 上聚合的背景流的平均强度.

由于 $\{p_1, m_1, q_1, m_2\}$ 和 $\{p_2, m_3, q_2, m_4\}$ 在到达 L_k 之前具有相同的分组结构并经过相同的网络环境, 可认为 $E(\Delta_{m,1}^k) = E(\Delta_{m,2}^k)$, 其中, $E(x)$ 是变量 x 的均值. 虽然 $\Delta_{m,1}^k$ 和 $\Delta_{m,2}^{k+1}$ 都会受到链路 $L_i (k < i \leq N)$ 的容量以及背景流的影响, 但由于 m_1 和 m_2 、 m_3 和 m_4 都是等长分组对, 所以只要 $\Delta_{m,1}^k$ 和 $\Delta_{m,2}^{k+1}$ 足够大, 这两个时间间隔可以保持到接收端, 即 $E(\Delta_{m,1}^k) = E(\Delta_{out,1}^N)$, $E(\Delta_{m,2}^{k+1}) = E(\Delta_{out,2}^N)$. 结合公式(1), 可以获得对链路 L_k 上背景流平均强度 λ_k 的估计, 进一步估计 L_k 上的可用带宽 A_k , 即

$$A_k = \frac{(E[\Delta_{out,1}^N] - E[\Delta_{out,2}^N])C_k + s(m) + s(q)}{E[\Delta_{out,1}^N]} \quad (2)$$

其中, C_k 是链路 L_k 的容量, 假设已知或由已有的链路容量测量工具获取. 不同于随机变化的可用带宽, 链路的容量一般是固定的. 在负载较低的情况下, 已有的链路容量测量方法能够取得较高的准确度. 实际上, LinkPPQ 经过稍加改变也可以用于对链路容量的测量.

2.2 四探测分组结构在被测量链路上与背景流在同一个缓冲区队列

通过设计四探测分组结构中分组大小, 使其经过被测量链路时与背景流在该链路的同一个缓冲区队列中.

为简化问题, 首先假设路径上没有背景流. 由于通常设置 $s(q) < s(p)$, 使得分组在传输过程中有靠拢的趋势, 当探针中 $\{p_1, m_1, q_1, m_2\}$ 和 $\{p_2, m_3, q_2, m_4\}$ 分别以一个紧挨着一个的形式进入 L_1 , 即当标志分组 m_1 (m_3) 和 m_2 (m_4) 进入 L_1 的间隔 $T_{intra} = [s(q) + s(m)] / C_1 \leq [s(p) + s(m)] / C_{1,k}$ 时, $\{p_2, m_3, q_2, m_4\}$ 能够一个紧挨着一个地离开 $B_{1,k}$ 中的任意一条狭窄链路.

考虑当 $L_k \notin B_{1,k}$ 时的情况. 在无背景流条件下, 当四探测分组结构 $\{p_2, m_3, q_2, m_4\}$ 一个紧挨着一个地离开 $B_{1,k}$ 中的任意一条狭窄链路时, 标志分组 m_3 与 m_4 的间隔达到最大值 $[s(q) + s(m)] / C_{1,k}$. 所以, 为保证 $\{p_2, m_3, q_2, m_4\}$ 保持一个紧挨着一个的形式离开 L_k , 其充分条件是

$$\frac{s(p_2)}{C_k} \geq \frac{s(m_3)}{C_{1,k}} \quad (3)$$

$$\frac{s(p_2) + s(m_3)}{C_k} \geq \frac{s(m_3) + s(q_2)}{C_{1,k}} \quad (4)$$

$$\frac{s(p_2) + s(m_3) + s(q_2)}{C_k} \geq \frac{s(m_3) + s(q_2) + s(m_4)}{C_{1,k}} \quad (5)$$

公式(3)~公式(5)说明,第1个探测分组 p_2 适合于采用大分组,即设置 $s(p)$ 尽可能地大.由于在以太网帧中数据的长度一般介于 46B 和 1500B 之间,因此,我们选择 $s(p)=1500B$.标志分组 m_3 和 m_4 的作用主要是分别跟随定速分组 p_2 和 q_2 ,其间隔反映定速分组的间隔.考虑到小分组在路径上传输所受的影响比较小,我们选择 $s(m)=46B$,标志分组之所以选择相同的长度,在第 2.3 节给出了解释.

我们重点选择 q_2 分组的大小.由于在 m_3 和 m_4 到达 L_k 的时间间隔 $\Delta_{m,2}^k$ 内到达 L_k 的背景流分组的数量与 $\Delta_{m,2}^k$ 成正比,所以, $s(q)$ 越大, $\Delta_{m,2}^k$ 也越大,从而探测分组能够捕捉更多关于背景流的信息.

定义:

$$\alpha_k = C_k / C_{1,k} \quad (6)$$

可知 $\alpha_k \geq 1$, α_k 不仅与链路的容量有关,还与链路的位置有关. $\alpha_k=1$ 表示 $L_k \in B_{1,k}$; $\alpha_k > 1$ 表示 L_k 的容量大于局部路径的容量,是非狭窄链路.假设 $s(q)$ 的上限用 $U(s(q))$ 表示.当 $\alpha_k=1$ 时, $U(s(q))$ 可取 1500B,这时,公式(3)~公式(5)都满足;当 $\alpha_k > 1$ 时,可根据公式(3)~公式(5)计算 $U(s(q))$.当 α_k 取整数时, α_k 与 $U(s(q))$ 的关系见表 1.

Table 1 Relationship between α_k and $U(s(q))$, given $s(p)=1500B$ and $s(m)=46B$

表 1 α_k 与 $U(s(q))$ 的关系,给定 $s(p)=1500B, s(m)=46B$

α_k	1	2	3	4	5	6	7	8	9	10	11
$U(s(q))$ (B)	1500	727	469.3	340.5	263.2	198.8	150.3	115.7	89.8	69.6	53.4

由于以太网帧的数据部分的最小分组长度一般为 46B,为保证 $U(s(q))$ 不小于 46B,要求 $\alpha_k \leq 11$,但这已经满足从容量为 10Mbps 的边界链路测量容量为 100Mbps 的主干链路的可用带宽,或者从容量为 100Mbps 的链路测量容量为 1Gbps 的链路的可用带宽的实际应用需要.

实际网络中一般存在背景流,下面我们分析局部路径 $P_{1,k}$ 上各链路背景流对探针的影响.由于标志分组比较小,跟随定速分组,所以标志分组与其前面的定速分组之间能够插入的背景流分组相对比较少.背景流的消极影响是插入在 $m_1(m_3)$ 与 $q_1(q_3)$ 之间的背景流分组增加了 $m_1(m_3)$ 与 $m_2(m_4)$ 之间的时间间隔(我们把 $q_1(q_2)$ 和 $m_2(m_4)$ 之间插入的背景流分组合加在 $q_1(q_2)$ 之前的背景流分组总长度之中);积极影响是先进入缓冲区队列的背景流分组使得 $p_1(p_2)$ 排队等待,从而使 $m_1(m_3)$ 与 $m_2(m_4)$ 之间的时间间隔可能缩小.最糟糕的情况出现在背景流使得 $m_1(m_3)$ 与 $m_2(m_4)$ 到达 L_k 的间隔过大,使得 $\{p_2, m_3\}$ 与 $\{q_2, m_4\}$ 经过 L_k 时不在同一个缓冲区队列.然而,LinkPPQ 探针的设计可以降低这些消极因素的影响.若 $\{p_2, m_3\}$ 与 $\{q_2, m_4\}$ 不在 $L_i(i < k)$ 的同一个缓冲区队列,则由文献[30]可知:

$$\Delta_{out,2}^i - \Delta_{in,2}^i = \frac{[s(q_2) + s(m_4)] - [s(p_1) + s(m_3)]}{C_i} + (\omega_{qm}^i - \omega_{pm}^i) = \frac{s(q) - s(p)}{C_i} + (\omega_{qm}^i - \omega_{pm}^i) \quad (7)$$

其中, ω_{pm}^i 和 ω_{qm}^i 分别是 $\{p_2, m_3\}$ 与 $\{q_2, m_4\}$ 的排队时延.由于 $\{p_2, m_3\}$ 与 $\{q_2, m_4\}$ 处于不同缓冲区队列, ω_{pm}^i 和 ω_{qm}^i 是独立同分布的随机变量,两者之差的均值等于 0.由于 $s(p) \geq s(q)$,那么 m_3 和 m_4 离开 L_i 的间隔 $\Delta_{out,2}^i$ 将在到达间隔 $\Delta_{in,2}^i$ 的基础上减去 $[s(p) - s(q)]/C_i$, $s(p)$ 和 $s(q)$ 差异越大,或 C_i 越小,则越有利于保证四探测分组结构 $\{p_2, m_3, q_2, m_4\}$ 在 L_k 的同一缓冲区队列中出现.

2.3 在路径上保持时间间隔

有必要保持探针 $\{p_1, m_1, q_1, m_2\}$ $\{p_2, m_3, q_2, m_4\}$ 中 m_1 与 m_2 到达被测量链路 L_k 的时间间隔 $\Delta_{m,1}^k$ 以及 m_3 与 m_4 离开 L_k 的时间间隔 $\Delta_{m,2}^{k+1}$ 到接收端.

分组对 $\{m_1, m_2\}$ 经过局部路径 $P_{i,j}$, 并且 m_1 和 m_2 都不在 $L_h (i \leq h \leq j)$ 的同一个缓冲区队列中, 那么, $\{m_1, m_2\}$ 离开 L_j 的间隔 Δ_{out}^j 与到达 L_i 的间隔 Δ_{in}^i 之差可由公式(7)相应地变换后累加获得:

$$\Delta_{out}^j - \Delta_{in}^i = \sum_{h=i}^j \frac{[s(m_2) - s(m_1)]}{C_h} + \sum_{h=i}^j (\omega_{m_2}^h - \omega_{m_1}^h) \quad (8)$$

由于 m_1 和 m_2 不同时出现在同一个缓冲区队列中, 所以排队时延 $\omega_{m_2}^h$ 和 $\omega_{m_1}^h$ 是独立同分布的随机变量, $\sum_{h=i}^j (\omega_{m_2}^h - \omega_{m_1}^h)$ 可认为是近似对称且均值为 0 的噪声. 设置 $s(m_2) = s(m_1) = s(m)$, 那么公式(8)等号右边第 1 项为 0, 于是 $E(\Delta_{out}^j) = E(\Delta_{in}^i)$, 从而达到保持分组间隔的目的. 这也是 LinkPPQ 将标志分组设置为等长分组的原因.

根据文献[28], 一般地, 在没有背景流的情况下, 等长分组对 $\{m_1, m_2\}$ 到达局部路径 $P_{i,j}$ 的间隔 Δ_{in}^i 能够保持到离开链路 L_j 的充分必要条件是

$$s(m) / \Delta_{in}^i \leq C_{i,j} \quad (9)$$

当局部路径 $P_{i,j}$ 存在背景流时, IGI/PTR^[7]表明, 当背景流是平稳的, 并且路径上只有一条同时是狭窄链路和承压链路的拥塞链路时, 如果分组对输入间隔 $\Delta_{in}^i \geq s(m) / A_{i,j}$, 那么分组对间隔在传输过程中将保持不变, 即 $E(\Delta_{out}^j) = E(\Delta_{in}^i)$; 否则, 分组对的输出间隔 Δ_{out}^j 将大于输入间隔 Δ_{in}^i . 然而, 当背景流是突发性的, 并且网络路径有多条拥塞链路时, 所需的输入间隔 Δ_{in}^i 要远大于 $s(m) / A_{i,j}$ 才能保持分组对间隔在传输过程中保持不变.

为了将探针 $\{p_1, m_1, q_1, m_2\} \{p_2, m_3, q_2, m_4\}$ 中 m_1 与 m_2 的时间间隔 $\Delta_{m,1}^k$ 和 m_3 与 m_4 的时间间隔 $\Delta_{m,2}^{k+1}$ 保持到接收端, 我们希望 $\Delta_{m,1}^k$ 和 $\Delta_{m,2}^{k+1}$ 足够大, 使得 $s(m) / \Delta_{m,1}^k$ 小于 $C_{k,N}$ 和 $A_{k,N}$, 并且 $s(m) / \Delta_{m,2}^{k+1}$ 小于 $C_{k+1,N}$ 和 $A_{k+1,N}$.

如第 2.2 节所述, 四探测分组结构 $\{p_2, m_3, q_2, m_4\}$ 被设计为在无背景流情况下以一个紧挨着一个的形式离开被测链路 L_k , m_3 和 m_4 离开 L_k 的最小时间间隔 $\min(\Delta_{m,2}^{k+1}) = [s(q) + s(m)] / C_k$, 即 $\max[s(m) / \Delta_{m,2}^{k+1}] = C_k / [1 + s(q) / s(m)]$. 类似地, 我们有 $\max[s(m) / \Delta_{m,1}^k] = C_{k-1} / [1 + s(q) / s(m)]$. 所以只要 $s(q) / s(m)$ 足够大, C_k 不远大于 $C_{k+1,N}$ 和 $A_{k+1,N}$, 且 C_{k-1} 不远大于 $C_{k,N}$ 和 $A_{k,N}$, 就有 $s(m) / \Delta_{m,2}^{k+1}$ 小于 $C_{k+1,N}$ 和 $A_{k+1,N}$, 以及 $s(m) / \Delta_{m,1}^k$ 小于 $C_{k,N}$ 和 $A_{k,N}$. 例如, 当 $\alpha_k = 1$ 时, $U(s(q)) = 1500B$, $s(q) / s(m) = 1500 / 46 = 32.61$, $\max[s(m) / \Delta_{m,2}^{k+1}] = C_k / 33.61$, $\max[s(m) / \Delta_{m,1}^k] = C_{k-1} / 33.61$; 当 $\alpha_k = 5$ 时, $U(s(q)) = 263.2B$, $\max[s(m) / \Delta_{m,2}^{k+1}] = C_k / 6.7$, $\max[s(m) / \Delta_{m,1}^k] = C_{k-1} / 6.7$. 显然, 在有背景流的情况下, 背景流对分组间隔将具有扩张作用, 上述条件更加容易得到满足.

由于探针 $\{p_1, m_1, q_1, m_2\} \{p_2, m_3, q_2, m_4\}$ 中标志分组设置得尽量小(46B), 意味着受到 L_k 之后链路容量及其上背景流的影响会比较小, 并且公式(2)采用标志分组到达接收端的时间间隔 $E(\Delta_{out,1}^N)$ 与 $E(\Delta_{out,2}^N)$ 之差来计算可用带宽, 因而可以进一步降低后续路径上背景流对标志分组时间间隔的影响.

2.4 LinkPPQ算法基本过程

LinkPPQ 算法的基本过程可分下面几个阶段:

- 获取被测链路 L_k 所在路径上各链路的容量. 由于链路容量一般是固定的, 通常假设链路的容量已知, 或者由现有的链路容量测量工具^[28-30]等测量获得;
- 针对被测链路 L_k 构造探针. 通过设置探针中各分组的地址和 TTL 域定位链路 L_k ; 根据公式(6)计算 α_k , 并且根据公式(3)~公式(5)计算 $s(q)$ 的上限 $U(s(q))$, 对其作适当调整后用作 $s(q)$ 的值; 设置所需的探针的平均速率以及数量;
- 发送端向接收端发送探测序列, 接收端接收标志分组, 并记录其到达的时间戳. 接收端完成接收过程后, 根据公式(2)计算 L_k 的可用带宽.

3 参数选择

LinkPPQ 对链路的可用带宽的测量主要受 3 个因素的影响: 一是路径上各链路的容量; 二是路径上各链路

的背景流的性质,包括强度以及平稳性等;最后一个算法本身.下面先构造一个仿真拓扑,通过改变链路的容量以模拟不同性质的网络路径,并且在仿真网络中引入实际的背景流,以便使仿真场景更接近于实际情况,然后探究 LinkPPQ 的参数选择:四探测分组结构中的 $s(q)$ 以及一次估计所需要的四探测分组结构数量.

3.1 仿真实验设置

在仿真环境下能够构造实际环境中难以实现的拓扑,并且灵活地引入背景流.我们使用网络研究领域广泛采用的 NS-2^[15]进行仿真.本文采用的拓扑结构类似于文献[12]的仿真拓扑配置,如图 3 所示.网络路径 P 由 7 条双向对称的链路 L_1, L_2, \dots, L_7 构成,链路的容量分别用 C_1, C_2, \dots, C_7 表示.仿真拓扑图中,主干路径 P 之外的链路的容量都设置为 100Mbps,并且是双向对称的.路由器采用先进先出的分组调度策略,而缓冲区管理采用弃尾机制.由于仿真拓扑中各节点的缓冲区采用默认的大小,没有设置成无限大,所以,当节点不能及时处理到达的分组时,会产生分组丢失的现象,符合实际网络中的情况.探测流由发送端 P_S 出发流向接收端 P_d .背景流 f_2, f_3, f_4, f_5, f_6 在网络路径 P 上仅经过一条链路(one-hop persistent),等效于经过相应链路上的汇聚流.节点 n_{2S} 和 n_{2d} 分别是经过链路 L_2 的背景流 f_2 的发送端和接收端.类似地, n_{3S}, n_{3d} 分别是背景流 f_3 的发送端和接收端.由于 L_1 和 L_7 分别连接发送端 P_S 和接收端 P_d ,通过测量流出 P_S 和流入 P_d 的分组数量,可以获得两条链路上的可用带宽.因此,仿真实验主要是测量局部路径 $P_{2,6}$ 上各链路 L_2, L_3, L_4, L_5, L_6 的可用带宽.

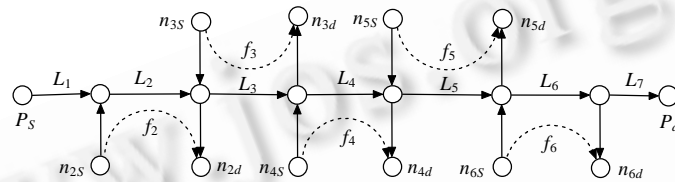


Fig.3 Simulation configuration

图 3 仿真拓扑配置

为使得仿真环境更贴近于实际,仿真中使用的背景流来自 WIDE 主干网^[32]的一条跨太平洋连接美国和日本容量为 150Mbps 的链路上若干天记录的采样文件,每天采样从中午 12:00 开始到 12:15 结束,持续 15 分钟.因此,仿真中产生的背景流分组长度的分布符合真实网络的情况.分组间隔被乘以一定的系数以产生符合要求的一定强度的背景流.仿真中所有的背景流的开始时间都是随机的,以避免时间同步.

3.2 性能评价标准

主要的测量评价标准有:

定义 1. 测量误差 ME(measurement error):假设链路或路径的实际可用带宽为 AV(actual value),测量得到的链路或路径的可用带宽为 MV(measured value),则测量误差 ME 定义为

$$ME = [(|MV - AV|) / AV] \times 100\% \quad (10)$$

定义 2. 测量平稳性 MS(measurement stability):假设对目标链路或路径进行了 n 次测量,测量得到的可用带宽值分别为 $V_1, V_2, \dots, V_n, \bar{V}$ 为测量值的平均值,则测量稳定性 MS 定义为

$$MS = \left[1 - \frac{1}{n} \sum_{i=1}^n (|V_i - \bar{V}| / \bar{V}) \right] \times 100\% \quad (11)$$

ME 考察 LinkPPQ 测量结果的准确程度,主要比较真实值与测量结果的平均值;而 MS 考察 LinkPPQ 测量结果的平稳性,反映多次测量结果之间的差异性.较低的 ME 和较高的 MS 表示测量结果比较好.

由于主动测量需要向网络中引入额外的探测流,为避免探测流增加网络的负担以及给网络中的应用造成严重影响,通过设置好如图 1 所示的四探测分组结构之间的间隔 T_{inter} ,使得四探测分组结构的平均发送速率为 500Kbps 或 5% 路径容量(取两者之中的最小值).

3.3 参数选择

LinkPPQ 主要有两个参数需要进行选择,其中一个是指针{p1,m1,q1,m2}{p2,m3,q2,m4}中 q1 和 q2 的大小 s(q), 另外一个获得一个被测链路的可用带宽的估计值所需的四探测分组结构的数量.

3.3.1 四探测分组结构中 s(q)的选择

我们选用 6 条典型的路径以模拟不同性质的网络拓扑,见表 2(单位:Mbps).每条路径应用 3 种不同的链路利用率组合以仿真实际网络中不同时段背景流强度的变化,如表 3 所示.其中的数值分别对应链路的利用率,由于 L1 和 L7 不是我们需要测量的链路,所以在上面没有额外引入背景流.从表 2 可以看出,所选的路径中有的含有一条狭窄链路,有的包括两条容量相同的狭窄链路;αk 的取值包括了从 1~10 的整数,以及部分介于 1 与 10 之间的非整数.当 αk 为整数时,对应的 U(s(q))在表 1 中.当 αk=1 时,s(q)的上限 U(s(q))=1500B;当 αk>1 时,由公式(3)~公式(5)可以获得 U(s(q)).于是,可以获得对表 2 中 6 条路径各链路进行测量时 s(q)的上限,见表 4(单位:B).在各链路对应的 U(s(q))的基础上,选用不同的 s(q),令 s(q)=βU(s(q)),其中,β的取值从 40% 开始以 10% 递增到 120%.考虑到以太网帧的数据部分的长度范围一般为[46B,1500B],故我们仅仿真调整后 s(q)在[46B,1500B]范围内的场景.

仿真的过程可描述如下:从表 2 选择一条路径 P,从表 3 中选用一组利用率,构造一个网络场景,共有 18 个场景.对于每个场景,选择被测量的链路 Lk,从表 4 中查到测量 Lk 时对应的 U(s(q)),然后把 s(q)设置为 βU(s(q)),当 s(q)在范围 [46B,1500B] 内时,发送一组足够数量(2 000)的四探测分组结构,即 1 000 个探针 {p1,m1,q1,m2}{p2,m3,q2, m4},获取 Lk 上可用带宽的一次估计值,隔一段时间,又发送一组四探测分组结构,再获得一个估计值,直到获得 25 个估计值,停止此 s(q)对应的测量;将β换一个值,重复对 Lk 的测量;当β遍历完从 40% 以 10% 递增到 120% 的值时,停止对 Lk 的测量,然后对此场景中其余链路作类似 Lk 的测量.最后分两种情况统计 LinkPPQ 在不同β下的测量性能.

Table 2 Paths for choosing parameters
表 2 用于参数选择的路径

Table with 8 columns: Path, L1, L2, L3, L4, L5, L6, L7. Rows P1 to P6.

Table 3 Utilization combinations applied to chosen paths
表 3 应用于所选择的路径上的利用率组合

Table with 6 columns: Utilization, L2, L3, L4, L5, L6. Rows A, B, C.

Table 4 U(s(q)) for measuring the links of the chosen paths

表 4 用于测量所选择的路径上各链路的 s(q)的上限 U(s(q))

Table with 6 columns: Path, L2, L3, L4, L5, L6. Rows P1 to P6.

图 4 显示,LinkPPQ 采用不同β测量αk=1 对应的链路的测量误差 ME 和测量平稳性 MS 的累积分布(αk=1, s(q)=βU(s(q))).图 4(a)和图 4(b)表明,除β=40% 以外,其余β对应的 ME 都低于 30%.随着β的增大,小于 10% 的 ME 有减少的趋势.正如第 2.2 节的分析,四探测分组结构在传输过程中每经过一条链路 Li,分组不等长设计将使得两标志分组的输出间隔在输入间隔的基础上减少[s(p)-s(q)]/Ci.显然,在 s(p)一定的情况下,s(q)越小,减少量越大,即使探测分组在到达被测链路之前经背景流较大,标志分组到达被测链路的间隔也能保证足够小,从而保证四探测分组结构能够同时出现在被测链路的同一个输出缓冲区队列中.然而,s(q)并不是越小越好,图 4(c)和图 4(d)表明,随着β的增大,大于 90% 的 MS 有增加的趋势.有几个原因可以解释较大的 s(q)导致较稳定的测量结果:较大的 s(q)意味着在被测链路的缓冲区队列中探测分组之间插入更多的背景流分组,从而捕捉更多的背

景流的信息;另外,由第 2.3 节的分析可知, $s(q)$ 越大,标志分组离开被测链路的间隔也越大,从而探测结果更不容易受被测链路之后的链路本身的容量以及其上的背景流的影响.一般地,实验结果表明,当 $\alpha_k=1$ 时, $U(s(q))=1500B$, β 取 80% 可以获得较好的测量性能.

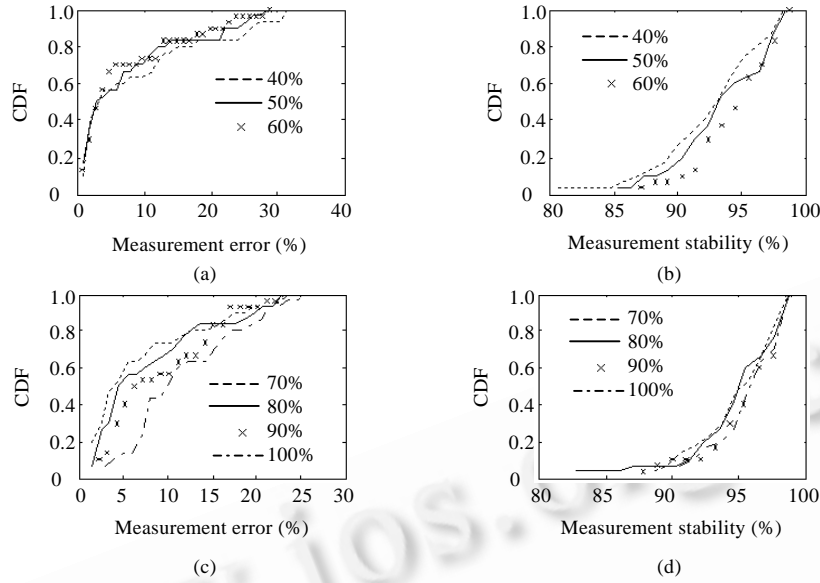


Fig.4 Cumulative distributions of measurement error and measurement stability with different β when $\alpha_k=1$

图 4 当 $\alpha_k=1$ 时,不同 β 对应的测量误差和测量平稳性的累积分布

图 5 显示,当 $\alpha_k>1$ 时,采用不同的 β 对应的测量误差 ME 和测量平稳性 MS 的累积分布 ($\alpha_k>1, s(q)=\beta U(s(q))$).

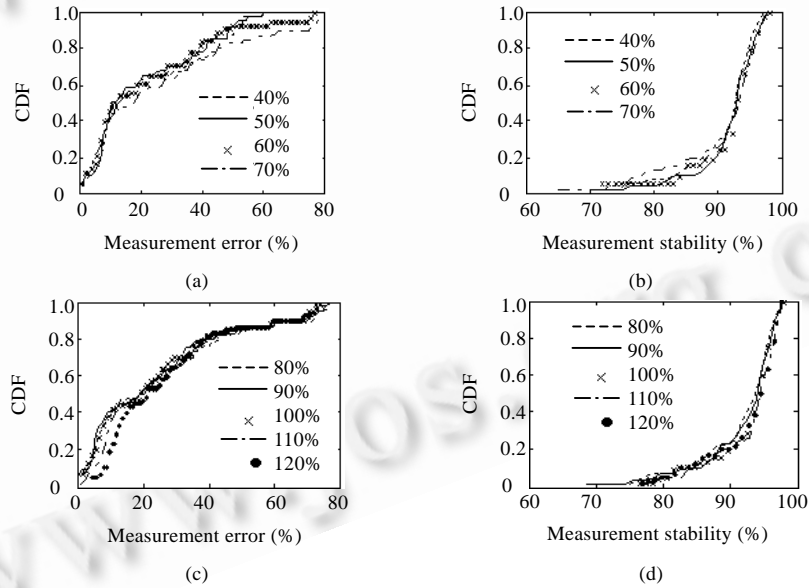


Fig.5 Cumulative distributions of measurement error and measurement stability with different β when $\alpha_k>1$

图 5 当 $\alpha_k>1$ 时,不同 β 对应的测量误差和测量平稳性的累积分布

当 β 为 40% 时,为避免 $s(q)$ 小于 $46B$, $U(s(q))$ 应不小于 $115B$, 所以 β 为 40% 对应的累积分布曲线没有包括 $\alpha_k=9$ 以及 $\alpha_k=10$ 的链路的仿真结果. 只有当 β 大于 $46/69=66.7\%$ 时, β 对应的曲线才包括所有 $\alpha_k>1$ 时链路的仿真结果.

经过统计,仿真实验中,大于 30%的 ME 中有 94%出现在对 $\alpha_k \geq 7$ 的链路的测量方面.这也是图 5(a)和图 5(b)中 β 为 40%或 50%时的测量性能比 β 为 60%或 70%时更好的原因.当 β 不小于 70%时,各累积分布曲线中小于 30%的 ME 占的比例最小,为 63.3%.而小于 10%的 ME 占的比例随着 β 的增大而减少;当 β 在 70%和 100%之间时,各累积分布曲线中大于 90%的 MS 的比例随着 β 的增大而增大.一般地,在 $\alpha_k > 1$ 时, β 取 100%时能够取得较好的性能.

我们在第 2.3 节的基础上分析 $\alpha_k \geq 7$ 时造成测量误差较大的原因.由表 1 所示,当 $\alpha_k \geq 7$ 时, $U(s(q)) < 151B$.探针 $\{p_1, m_1, q_1, m_2\} \{p_2, m_3, q_2, m_4\}$ 中标志分组 m_1 和 m_2 进入 L_k 的间隔的最小值:

$$\min(\Delta_{m,1}^k) = [s(q) + s(m)] / C_{k-1} \leq [U(s(q)) + s(m)] / C_{k-1} < [1 + 151/s(m)] \times s(m) / C_{k-1} = 4.28 \times s(m) / C_{k-1}.$$

在表 2 中, P_2, P_4, P_5 的部分链路的 α_k 大于或等于 7.在测量 P_4 的 L_5 时, $\alpha_5 = 8$,对应的 $U(s(q)) = 115B$.当 $s(q)$ 设置为 $U(s(q))$ 时,标志分组 m_1 和 m_2 进入 L_5 的最小间隔为 $\min(\Delta_{m,1}^5) = [s(q) + s(m)] / C_4 = 4.28 \times s(m) / 60Mbps$,小于 m_1 和 m_2 在 P_4 上的一条狭窄链路 L_6 的最小间隔 $s(m) / 10Mbps$,所以, m_1 和 m_2 进入 L_5 的间隔 $\Delta_{m,1}^5$ 很可能在 L_6 发生较大的改变.类似地,由于 m_3 和 m_4 离开 L_5 的时间间隔 $\Delta_{out,2}^5$ 较小,也可能被 L_6 本身的容量以及背景流严重影响.由于部分 $\Delta_{m,1}^5$ 和 $\Delta_{out,2}^5$ 不能保持到接收端,以及 L_5 之前链路的背景流较大可能造成 $\{p_2, m_3, q_2, m_4\}$ 不能在 L_5 的同一个缓冲区队列中出现,这些因素均造成测量出现较大误差.然而,只要 α_k 不要过大,相应地 $s(q)$ 也较大,当局部路径 $P_{1,k-1}$ 的可用带宽和 L_k 的可用带宽相比不是很小,并且 $P_{k+1,N}$ 的容量和可用带宽不远小于 L_k 以及前面链路的容量和可用带宽时,那么 LinkPPQ 会有较好的测量性能.

综上所述,当所测链路对应的 $\alpha_k = 1$ 时, β 取 80%,即 $s(q) = \beta U(s(q)) = 80\% \times 1500B = 1200B$;当 $\alpha_k > 1$ 时, β 取 100%,由于存在 $U(s(q)) > 1200B$ 的情况,所以选择 $s(q) = \min(1200B, U(s(q)))$.

3.3.2 四探测分组结构数量的选择

仿真环境与上节一样,所不同的是 $s(q)$ 的大小根据上一节总结的值进行选取.实际上,只要将表 4 中的 1 500B 替换为 1 200B;在上节中进行一次估计的四探测分组结构的数量统一设置为 2 000 个,在这里,用于一次估计的四探测分组结构的数量在范围[100,1500]内,数量以 100 递增.

仿真结果表明,当四探测分组结构的数量在 100 以上时,数量的改变对测量误差没有很明显的影.数量与测量平稳性 MS 的关系如图 6 所示.从图中可以看出,当数量在 500 以上时,随着数量的增多,MS 的均值增大不是很显著;同样地,MS 的均方差降低得也不多.图 6 左图表示 $\alpha_k = 1$ 时的 MS 的均值比 $\alpha_k > 1$ 时高 2.46%左右;而图 6 右图显示 $\alpha_k = 1$ 对应的 MS 的均方差比 $\alpha_k > 1$ 时的小 1.50%左右.

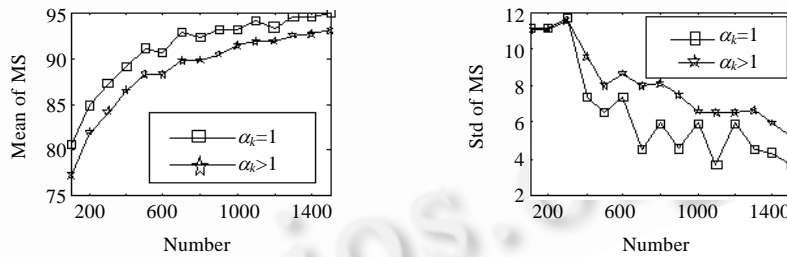


Fig.6 Number of packet-quartets vs. mean and std of MS

图 6 MS 的均值和均方差与四探测分组结构数量的关系

进行一次测量所需的从发送端引入的负载 MO(measurement load)和所需的时间 t 与四探测分组结构数量 num 有紧密的关系.假设四探测分组结构的发送速率为 r ,则 $MO = [s(p) + s(q) + 2 \times s(m)] \times num$;而 $t = MO / r$.显然,MO 和 t 都正比于 num .图 6 表明, num 大于 500 后,随着 num 的增大,测量性能没有很明显的变化,但同时 MO 和 t 都线性增大.同时考虑到测量性能和测量负载, $num = 500$ 是合理的选择.

3.3.3 LinkPPQ 对网络的影响

一般地,四探测分组结构的发送速率 r 设置为 500kbps 或路径容量的 5%,取两者中的最小值.在图 1 中,相邻

四探测分组结构的时间间隔 $T_{inter}=[s(p)+s(q)+2 \times s(m)]/r$. 设定一次测量所需的四探测分组结构数量 num 为 500. 由于 $s(q) \leq 1200B$, 那么一次测量的从发送端引入的负载为

$$MO=[s(p)+s(q)+2 \times s(m)] \times num \leq [1500+1200+46+46] \times 8 \times 500 \times 10^{-6} = 11.17Mb.$$

而 $t=MO/r=11.17Mb/500kbps=22.34s$, 此时 $r=500kbps$. 由于探针 $\{p_1, m_1, q_1, m_2\}$ $\{p_2, m_3, q_2, m_4\}$ 中 p_1 和 q_1 在到达 L_k 就因 TTL 超时而被丢弃, 而 p_2 和 q_2 在离开 L_k 时也因 TTL 超时而被丢弃, 所以一次测量中, 四探测分组结构在局部路径 $P_{1,k-1}$ 的平均速率为 r , 在 L_k 上的平均速率降到 $r \times [1-0.5/[1+2 \times s(m)/(s(q)+s(p))]]$, 约等于 $0.5r$, 而在局部路径 $P_{k+1,N}$ 上的平均速率降到 $r \times [2 \times s(m)]/[s(q)+s(p)+2 \times s(m)]$. 这样减轻了对被测链路 L_k 和局部路径 $P_{k+1,N}$ 的影响, 而目前路径可用带宽测量工具带给路径上所有的链路相同的负载. 由于实际网络中背景流至少在数分钟内是平稳的^[1], 而 LinkPPQ 进行一次测量的时间小于 30s, 所以 LinkPPQ 具备实时监控链路的可用带宽的能力.

4 实验验证

4.1 仿真验证

我们选用两种类型的路径: 具有单狭窄链路的路径和具有多狭窄链路的路径, 然后引入背景流, 构造多个不同的背景流场景. 我们选用路径 $P_m^1 = \{C_1, C_2, \dots, C_7\} = \{100, 30, 60, 10, 60, 20, 100\}$ Mbps, 其中 L_4 是狭窄链路. 测量结果表明, 一般地, 测量 $\alpha_k=1$ 时对应链路的性能好于 $\alpha_k > 1$ 的, 但是无论是对承压链路或其前后链路的测量, 大多数 ME 均小于 30%, 而 MS 均高于 80%. 我们这里重点描述对具有多狭窄链路的路径的测量结果.

具有多狭窄链路(multiple narrow link)的路径 $P_{mm}^1 = \{C_1, C_2, \dots, C_7\} = \{100, 10, 40, 20, 30, 10, 100\}$, 应用 7 种利用率组合构成 7 个测量场景, 其中, 4 种利用率组合见表 5, 另外 3 种为 $UC_1=0.1, UC_2=0.3, UC_3=0.5$, 分别表示 $L_2 \sim L_6$ 上的利用率都为 0.1, 0.3 或 0.5. 测量 $L_2 \sim L_6$ 的 $s(q)$ 分别设置为 1 200B, 340B, 727B, 469B, 1 200B, 获得一次估计值的探测序列中四探测分组结构的数量设置为 500. L_2 和 L_6 是 P_{mm}^1 上的两条狭窄链路, 当应用 UC_1, UC_2, UC_3 时, L_2 和 L_6 同时是承压链路; 当应用 UC_4 和 UC_5 时, L_2 是承压链路; 当应用 UC_6 和 UC_7 时, L_6 是承压链路. 这在一定程度上反映了背景流对承压链路的位置会有影响. 测量 L_2, L_3, \dots, L_6 的结果见表 6. 从表 6 可以看出, 在各个场景, 测量狭窄链路 L_2 的性能都好于测量狭窄链路 L_6 的性能, 前者 ME 都不大于 6.16%, 后者 ME 不大于 25.75%, 两者的 MS 都在 88% 以上. 之所以出现性能方面的差异, 主要是因为测量 L_2 时, L_1 上没有背景流, 从而不会影响探测分组. 另外, 标志分组到达 L_2 与离开 L_2 的间隔足够大, 受到 $L_3 \sim L_6$ 的链路容量和背景流的影响较小; 而测量 L_6 时, 探测分组会受到从 $L_2 \sim L_5$ 的链路容量和背景流的影响, 可能使得标志分组进入 L_6 时的间隔较大, 从而一定程度上影响探测分组与 L_6 上背景流在同一个缓冲区队列排队. 然而探测分组的不对称设计, 一定程度地降低了进入 L_6 之前的链路造成的影响, 使得测量 L_6 的 ME 都不大于 25.75%. 另外, 由于测量 L_2 和 L_6 的 $s(q)$ 为 1 200B, 从而标志分组之间间隔能够捕获较多的背景流信息, 使得测量结果比较平稳, MS 都在 88% 以上.

非狭窄链路 L_3, L_4, L_5 对应的 $\alpha_3=4, \alpha_4=2, \alpha_5=3$. 测量这些链路时, 除个别情况外, ME 都在 30% 以下, MS 都在 80% 以上. 表 6 也表明, 随着 $L_2 \sim L_6$ 上的利用率由 UC_1 增大到 UC_3 , ME 有增加的趋势, 而 MS 有降低的趋势, 说明被测链路前后各链路的容量和背景流对测量施加了不利的影响. LinkPPQ 通过设计特殊的分组结构, 尽量避免背景流和容量的负面影响.

Table 5 Different utilization combinations (UCs) used in simulation validation

表 5 仿真验证的不同利用率组合

UC	L_2	L_3	L_4	L_5	L_6
UC_4	0.4	0.1	0.2	0.3	0.3
UC_5	0.5	0.4	0.3	0.5	0.3
UC_6	0.4	0.3	0.4	0.1	0.5
UC_7	0.1	0.5	0.3	0.6	0.2

一般地, 在测量拥有多狭窄链路的路径时, 测量靠近发送端的狭窄链路的性能要好于测量离发送端较远的狭窄链路的性能. LinkPPQ 同样可以对狭窄链路前后的非狭窄链路进行有效的测量, 大多数 ME 都在 30% 以下, 而 MS 大多数都在 80% 以上.

Table 6 Measurement error and measurement stability when LinkPPQ measures all links except L_1 and L_7 on the path with multiple narrow links under several cross-traffic scenarios

表 6 LinkPPQ 对处于不同背景流场景下具有多狭窄链路的路径上除 L_1 和 L_7 外所有链路进行测量的误差和平稳性

Utilization		$UC_1=0.1$	$UC_2=0.3$	$UC_3=0.5$	UC_4	UC_5	UC_6	UC_7
ME (%)	L_2	1.87	4.09	6.16	1.52	4.55	4.22	4.81
	L_3	10.11	17.86	20.09	26.62	24.45	26.07	11.20
	L_4	11.45	17.91	27.38	22.89	29.87	24.14	17.32
	L_5	10.15	22.10	30.28	17.97	24.15	31.46	16.12
	L_6	6.87	14.84	25.75	15.68	23.05	18.95	15.64
MS (%)	L_2	97.61	95.43	90.31	93.37	88.78	92.59	96.17
	L_3	95.29	90.64	83.37	93.07	86.76	88.64	82.51
	L_4	95.24	93.20	88.50	91.97	93.42	86.55	94.10
	L_5	94.74	89.65	82.05	88.60	83.64	89.18	70.76
	L_6	97.95	94.95	90.81	96.69	96.65	91.08	95.94

4.2 实验网验证

我们搭建了一个实验网,拓扑结构类似于图 3,采用和仿真实验中相同的背景流量文件.Tcpreplay^[33]用于读取实际的背景流记录并产生符合真实网络中背景流性质的背景流.在理论分析阶段,为简化问题,我们假设背景流是平稳的,但在仿真实验和本实验网中产生的背景流显然是突发性的,更符合真实网络的情况.实验网中的各中间节点都是传统的路由器,而不是用普通 PC 模拟而成的.Net-snmp^[3]运行在一台 Redhat Linux 工作站上,每隔 1 分钟访问实验网上路由器的 MIB 库获得流过相应链路的流量强度,再减去探测流在相应链路的强度,获得背景流强度.链路的容量减去背景流强度,作为相应链路的真实可用带宽.值得指出的是,计算真实背景流强度时不包括探测流强度,探测流在相应链路的速率可参照第 3.3.3 节中的计算.

我们在 3 种不同类型的路径上进行了测量.

4.2.1 由低容量链路测量高容量链路的可用带宽

路径 $P_{LH}=(10,100,100)$ Mbps.主要测量链路 L_2 .每次获得一个估计值,发送 600 个四探测分组结构,其中, $s(q)$ 设置为 69B.6 次改变 L_2 上的背景流强度,构造 6 个不同的网络场景.测量结果见表 7.表中 Net-snmp 对应真实的可用带宽值.Duration 表示进行一次简单的估计所需要的时间.Lossrate 表示分组丢失率,把四探测分组结构中的标志分组对作为统计对象,如果标志分组中任意一个或两个都丢失,表示此标志分组对丢失,丢失的数量除以四探测分组结构的数量,得到丢失率.丢失率能够反映路径上的真实拥塞程度,能够帮助了解测量误差增加的原因.从表 7 中可以看出,LinkPPQ 在各个场景下测量误差都小于 20%,单次测量的时间的平均值都小于 22s.随着背景流强度的增大,背景流的突发性更强,表现为 Net-snmp 可用带宽测量值的均方差逐渐增大.Lossrate 由低强度背景流场景下的 1.53% 增加到 10% 以上,LinkPPQ 测量值的均方差也逐渐增大,单次测量的时间也有所增加,反映了背景流的强度和突发性对测量性能的影响.总体来说,从容量为 10Mbps 的链路来测量容量为 100Mbps 链路的可用带宽,LinkPPQ 具有良好的测量性能,显著不同于现有的路径可用带宽测量算法.

Table 7 When measuring L_2 on $P_{LH}=(10,100,100)$ Mbps, LinkPPQ performance vs. the true values measured by Net-snmp

表 7 测量路径 $P_{LH}=(10,100,100)$ Mbps 中 L_2 ,比较 LinkPPQ 测量性能与 Net-snmp 采集的真实值

LinkPPQ				Net-snmp		ME (%)
Mean (Mbps)	Std (Mbps)	Lossrate (%)	Duration (s)	Mean (Mbps)	Std (Mbps)	
82.837 5	4.077	1.53	16.119 7	89.505 9	0.008 4	7.450 2
71.462 1	9.690 8	13.04	17.616 9	78.984 2	0.02	9.523 5
65.983 2	12.154 9	15.8	18.018 8	68.508 6	0.200 4	3.686 2
61.833 5	10.040 4	14.26	19.243 7	59.066 8	2.589 7	4.683 9
56.425 4	12.571 1	12.32	20.828 8	50.347 9	4.708 9	12.070 9
52.118	7.325 2	12.08	21.533 3	44.300 9	7.240 1	17.645 6

4.2.2 具有单狭窄链路的路径

路径 $P_{sm}^2=(100,100,10,100)$ Mbps,具有一条狭窄链路 L_3 .测量 L_2 ,考察对狭窄链路之前链路的可用带宽进行测量的性能,并且观察狭窄链路的容量与背景流对测量结果的影响.测量 L_2 的结果如图 7 所示.

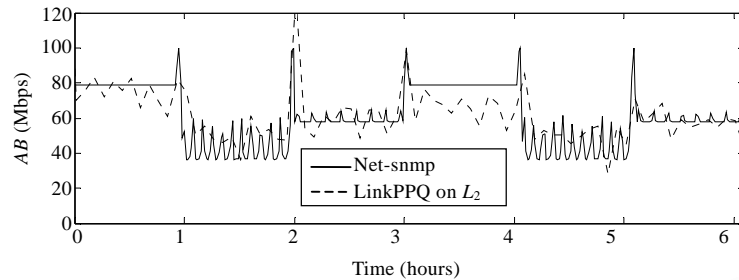


Fig.7 When measuring L_2 on $P_{sm}^2=(100,100,10,100)$ Mbps, LinkPPQ vs. Net-snmip

图 7 测量 $P_{sm}^2=(100,100,10,100)$ Mbps 上的 L_2 时,比较 LinkPPQ 测量结果和 Net-snmip 的真实值

在时间段[0,3]小时和(3,6)小时, L_3 上背景流平均强度分别为 2.25Mbps 和 4.07Mbps.在每时间段内, L_3 的背景流比较平稳,而 L_2 上背景流平均强度每小时改变一次,各小时时间段内背景流的平均值分别为 21.02Mbps, 40.94Mbps 和 55.69Mbps.为获取一个 L_2 的可用带宽估计值,发送端发送 600 个四探测分组结构,其中, $s(q)$ 设置为 1 350B.在每次估计结束后、新的估计开始前,停止 5s,使得网络有足够的时间恢复到没有探测流的状态.图 7 的曲线是每 5 分钟取一次平均值而获得的.图中,Net-snmip 对应的曲线的变化能够体现背景流的突发性,背景流的平均强度越大,突发性显得越强.虽然在时间段[0,1]小时和(3,4)小时, L_3 上背景流的改变使得测量结果有些变化,但是随着 L_2 背景流的增大,从时间段(1,3]小时和(4,6]小时的曲线可以看出, L_3 上背景流平均强度的增大,对 L_2 的测量结果影响很小,即探针 $\{p_1, m_1, q_1, m_2\}$ 和 $\{p_2, m_3, q_2, m_4\}$ 中标志分组 m_1 和 m_2 进入 L_2 的间隔 $\Delta_{m,1}^2$ 以及 m_3 和 m_4 离开 L_2 的间隔 $\Delta_{m,2}^2$ 受 L_3 的影响比较小.Lossrate 的分布为:在区间[0,1]小时,Lossrate 平均为 13%,而在区间(3,4)小时,Lossrate 平均为 14%左右;在(1,2)小时和(4,5)小时,Lossrate 在 12%附近很小地波动;在区间(2,3)小时和(5,6)小时,Lossrate 介于 14%和 16%之间.这说明,随着 L_3 的背景流的增大,除了区间(3,4)外,Lossrate 并没有明显的增大,一定程度上反映出链路 L_2 是标志分组丢失的主要地方,当 L_2 的背景流在 20Mbps 以上时,丢失率没有随着背景流的增大而增大,而保持在 10% 以上的一个范围内.这与表 7 显示的丢失率的趋势是一致的.测量时间每 5 分钟的平均值绝大多数在 [28,32]s 之间,测量时间的变化主要是由于探测分组的发送过程结束后,发送端采用 UDP 分组向接收端发送命令,告诉接收端计算可用带宽,并把结果传回给发送端,在一定时间内,如果没有得到对方的响应,那么重新发送 UDP 分组,此过程最多重复 10 次.由于等待响应时间设置为 3s,测量时间说明大多数情况下,只要重复一次命令,传输过程就能获取可用带宽值.

总体上,对于具有单狭窄链路的路径,LinkPPQ 能够比较准确地测量狭窄链路之前的链路的可用带宽,实时跟踪背景流的变化,并且狭窄链路上背景流在一定范围内的增大对测量结果影响较小.

4.2.3 具有多狭窄链路的路径

路径 $P_{mm}^2=(10,10,100,100,10,10)$ Mbps,具有 4 条容量相同的狭窄链路,分别为 L_1, L_2, L_5, L_6 .我们主要测量 L_2 和 L_5 .容量为 100Mbps 的链路 L_3 和 L_4 上的背景流在整个测量过程中平均强度在 10Mbps 附近.

L_2 的背景流在时间段[0,62]分钟的背景流强度大约为 2Mbps,在时间段(63,125)分钟的利用率为 0.4,在(127,180)分钟背景流的平均强度降到 0.2Mbps,在时间段(62,63)和(125,126)分钟, L_2 上无背景流. L_5 上的背景流的平均强度每 20 分钟变化一次,每次变化之间有 1 分钟无背景流,背景流的平均强度在 2Mbps,6Mbps,4Mbps 这 3 个值循环.测量 L_2 时,每次测量,LinkPPQ 发送 500 个四探测分组结构,其中, $s(q)$ 设置为 1 200B;测量 L_5 时,每次测量,LinkPPQ 发送 600 个四探测分组结构,其中, $s(q)$ 设置为 1 050B.LinkPPQ 循环测量 L_2 和 L_5 ,获得 L_2 的可用带宽的一个估计值所需的平均测量时间为 22s,而测量 L_5 所需平均时间为 25s,LinkPPQ 在测量 L_2 和 L_5 的

间隙中,停止 13s.这样,每分钟 LinkPPQ 可获得 L_2 和 L_5 上各一次可用带宽估计值.

LinkPPQ 测量 L_2 和 L_5 的结果与 Net-snmp 测量的真实值的比较如图 8 所示.图 8(a)表明,无论 L_5 上背景流强度在 2Mbps 和 6Mbps 之间如何变化,LinkPPQ 都能较准确地测量 L_2 的可用带宽.这说明探针中的标志分组离开 L_2 后没有受后面的链路较大的影响,显示了良好的健壮性.当 L_2 的背景流强度约为 2Mbps 时,LinkPPQ 的测量曲线很接近 Net-snmp 的测量曲线(标识为 Net-snmp);当 L_2 的背景流强度约为 4Mbps 时,LinkPPQ 高估可用带宽 1Mbps 左右,测量误差 ME 约为 16.67%.图 8(b)显示,当 L_5 的背景流强度大于 L_2 的背景流强度时,LinkPPQ 能够准确地测量 L_5 的可用带宽;当前者等于后者时,LinkPPQ 的测量误差在 30% 范围内;当前者小于后者时,如时间段(63,83)分钟,LinkPPQ 的误差在 30% 附近,但大于 30%.当 L_2 和 L_5 的背景流强度相等时,LinkPPQ 测量 L_5 的 ME 小于 30%,主要是因为四探测分组结构的特殊结构使得探测分组在传输过程中出现靠近的趋势;然而,当 L_2 的背景流强度大于 L_5 时,由于 L_2 较大地增加了四探测分组结构中标志分组之间的间隔,从而使得四探测分组结构在 L_5 与背景流可能不能在同一个缓冲区队列中,即标志分组的间隔与背景流强度没有呈现线性关系,所以可用带宽被低估.

LinkPPQ 主要针对链路进行测量.路径上链路的可用带宽的最小值也就是路径的可用带宽,所以,LinkPPQ 也能测量路径可用带宽.

总体上,对于具有多狭窄链路的路径,当 LinkPPQ 测量狭窄链路的可用带宽时,被测量链路后面的链路对测量结果影响较少;测量离发送端较近的狭窄链路(非 L_1),LinkPPQ 的性能比较好;当被测量的狭窄链路的背景流大于前面狭窄链路的背景流时,LinkPPQ 的测量误差比较小.

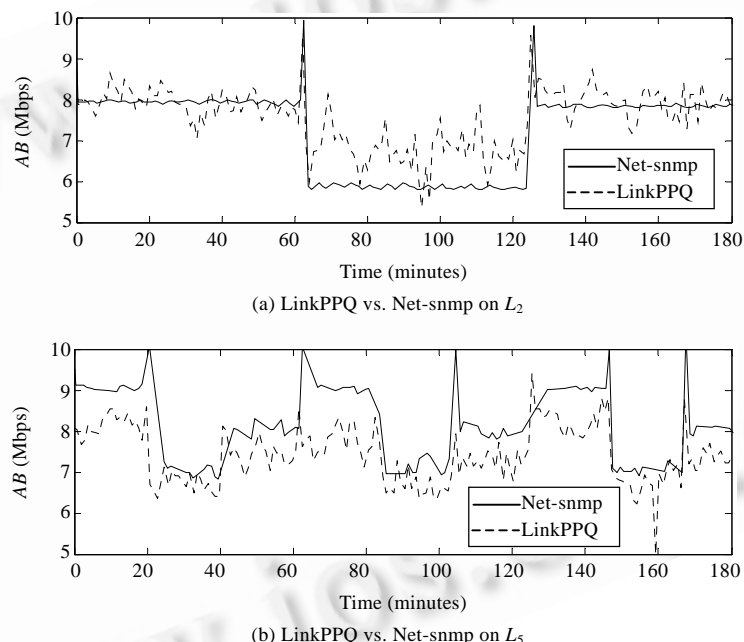


Fig.8 When measuring L_2 and L_5 on $P_{mm}^2=(10,10,100,100,10,10)$ Mbps, LinkPPQ vs. Net-snmp

图 8 测量 $P_{mm}^2=(10,10,100,100,10,10)$ Mbps 上 L_2 和 L_5 时,比较 LinkPPQ 估计值和 Net-snmp 的真实值

4.2.4 实验环境和程序设计对算法的影响

由于实验中需要精确地设置发送分组的间隔以及记录标志分组到达接收端的时间,并且在计算可用带宽之前需缓存这些数据,发送端和接收端都配置有 2.4GHz 的 CPU,512M 的内存,运行 Redhat Linux 9.0 操作系统,并且记录时间的精度为 $1\mu\text{s}$.LinkPPQ 采用 UDP 协议传输探针和控制信息,避免实际测量时由于路由器的背景流强度较大和突发性较高,出现无法建立 TCP 连接的情况.通过采用 UDP 分组多次重传控制信息的方式,来较

高程度地保证发送端和接收端交换控制信息.

分组的丢失也对目前的路径可用带宽算法造成影响.由于部分探测分组被中间路由器丢失,为保证测量结果的平稳性,需要发送更多的探测分组,这无疑增加了网络的开销.

如果源和接收端之间存在多条路径,由于网络中采用类似负载均衡的策略,那么探测分组不能保证通过同一条路径,此时,本文的 LinkPPQ 和文献已有的路径可用带宽算法都无法对路径进行有效的测量.

由于实际环境与仿真环境的差异,LinkPPQ 在仿真获得的经验参数的基础上有些改变,主要是获取每次测量值的探测分组由 500 改为 600,测量 $\alpha_k=1$ 的链路的 $s(q)$ 在 1 200B 的基础上有些改变,当 L_k 靠近发送端时, $s(q)$ 稍大些,而靠近接收端时, $s(q)$ 稍小些.

5 结 论

本文提出了一种测量任意链路的可用带宽的算法 LinkPPQ,它能够定位测量网络中任意链路,包括测量此链路的可用带宽与跟踪此链路的背景流的变化.本文首先介绍了为本算法特殊设计的探针结构,从理论上分析了这种结构所具备的对任意链路的可用带宽进行测量的能力,研究了探测分组大小以及探测序列长度对 LinkPPQ 性能的影响;仿真验证了 LinkPPQ 具备对单狭窄链路路径和多狭窄链路路径上的各链路的可用带宽进行准确、平稳测量的能力.实际网络实验进一步验证了 LinkPPQ 能够准确地测量 10Mbps 链路之前或之后的 100Mbps 链路的可用带宽,并且其测量结果受该 10Mbps 链路负载的影响较小,这使得 LinkPPQ 显著不同于目前的路径可用带宽测量算法;对于具有多狭窄链路的路径,LinkPPQ 仍然能够比较准确地测量各狭窄链路的可用带宽.

References:

- [1] Zhang Y, Duffield N, Paxson V, Shenker S. On the constancy of Internet path properties. In: Proc. of the IMC. San Francisco, 2001. 197–211. <http://portal.acm.org/citation.cfm?doid=505202.505228>
- [2] Paxson V. End-to-End routing behavior in the Internet. IEEE/ACM Trans. on Networking, 1997,5(5):601–615.
- [3] Net-snmp. <http://net-snmp.sourceforge.net/>
- [4] cflow. <http://www.gnu.org/software/cflow/>
- [5] sflow. <http://www.sflow.org/>
- [6] NetFlow. <http://www.cisco.com/go/netflow/>
- [7] Hu N, Steenkiste P. Evaluation and characterization of available bandwidth probing techniques. IEEE Journal on Selected Areas in Communications, 2003,21(6):879–894.
- [8] Navratil J, Cottrell RL. ABWE: A practical approach to available bandwidth estimation. In: Proc. of the Passive and Active Measurement Workshop. La Jolla, 2003. <http://www-iepm.slac.stanford.edu/tools/abing/abwe-pam03-paper.pdf>
- [9] Ribeiro V, Coates M, Riedi R, Sarvotham S, Hendricks B, Baraniuk R. Multifractal cross-traffic estimation. In: Proc. of the ITC Specialist Seminar on IP Traffic Measurement, Modeling and Management. Monterey, 2000. http://www.stat.rice.edu/~vinay/talks/itc00_talk.ps.gz
- [10] Ribeiro VJ, Riedi RH, Baraniuk RG, Navratil J, Cottrell L. pathChirp: Efficient available bandwidth estimation for network paths. In: Proc. of the PAM. San Diego, 2003. 200–211. <http://www-ece.rice.edu/~vinay/papers/pam03.pdf>
- [11] Strauss J, Katabi D, Kaashoek F. A measurement study of available bandwidth estimation tools. In: Proc. of the IMC. Miami, 2003. 39–44. <http://www.icir.org/vern/imc-2003/papers/p332-strauss.pdf>
- [12] Jain M, Dovrolis C. End-to-End available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput. IEEE/ACM Trans. on Networking, 2003,11(4):537–549.
- [13] Melander B, Bjorkman M, Gunningberg P. Regression-Based available bandwidth measurements. In: Proc. of the SPECTS. San Diego, 2002. <http://www.sics.se/cna/connected/spects2002.pdf>
- [14] Sommers J, Barford P, Willinger W. Laboratory-Based calibration of available bandwidth estimation tools. Microprocessors and Microsystems, 2007,31(4):222–235.
- [15] Ns2: Network simulator. http://nsnam.isi.edu/nsnam/index.php/Main_Page

- [16] Prasad R, Dovrolis C, Murray M, Claffy K. Bandwidth estimation: Metrics, measurement techniques, and tools. *IEEE Network*, 2003,17(6):27–35.
- [17] Keshav S. A control-theoretic approach to flow control. In: *Proc. of the ACM SIGCOMM'91*, Vol.21. Zurich, 1991. 3–15. <https://eprints.kfupm.edu.sa/17135/1/17135.pdf>
- [18] Carter R, Crovella M. Measuring bottleneck link speed in packet-switched networks. *Performance Evaluation*, 1996,27-28: 297–318.
- [19] Network characterization service (NCS). <http://www-didc.lbl.gov/NCS/>
- [20] Dovrolis C, Ramanathan P, Moore D. What do packet dispersion techniques measure? In: *Proc. of the IEEE INFOCOM*. Anchorage, 2001. 905–914. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=916282
- [21] Antoniadou D, Athanatos M, Papadogiannakis A, Markatos EP, Dovrolis C. Available bandwidth measurement as simple as running wget. In: *Proc. of the PAM*. Adelaide, 2006. <http://www.ics.forth.gr/~papadog/abget/abget.pdf>
- [22] Lv SH, Cai ZP, Yin JP, Zhao WT. Towards estimating available bandwidth on Network paths with application requirements. *Chinese Journal of Electronics*, 2006,34(10):1793–1798 (in English with Chinese abstract).
- [23] Liu M, Li ZC, Guo XB, Deng H. An end-to-end available bandwidth estimation methodology. *Journal of Software*, 2006,17(1): 108–116 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/17/108.htm>
- [24] Liu X, He L, Yu S. Algorithms of accurate network available bandwidth measurement. *Acta Electronica Sinica*, 2007,35(1):68–72 (in Chinese with English abstract).
- [25] Akella A, Seshan S, Shaikh A. An empirical evaluation of wide-area Internet bottlenecks. In: *Proc. of the IMC*. Miami, 2003. 101–114. <http://portal.acm.org/citation.cfm?id=948219>
- [26] Hu N, Li LE, Mao ZM, Steenkiste P, Wang J. Locating Internet bottlenecks: Algorithms, measurements, and implications. In: *Proc. of the ACM SIGCOMM*. Portland, 2004. <http://www.cs.cmu.edu/~hnn/papers/bottleneck.pdf>
- [27] Ribeiro VJ, Riedi RH, Baraniuk RG. Spatio-Temporal available bandwidth estimation with STAB. *IEEE Internet Computing Magazine*, 2004,8(5):34–41.
- [28] Harfoush K, Bestavros A, Byers J. Measuring bottleneck bandwidth of targeted path segments. In: *Proc. of the IEEE INFOCOM*. San Francisco, 2003. 2079–2089. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1209229
- [29] Lai K, Baker M. Measuring link bandwidths using a deterministic model of packet delay. In: *Proc. of the ACM SIGCOMM*. Stockholm, 2000. 283–294. http://www.hpl.hp.com/personal/Kevin_Lai/projects/nettimer/publications/sigcomm2000/deterministic.pdf
- [30] Pasztor A, Veitch D. Active probing using packet quartets. In: *Proc. of the IMC*. Marseille, 2002. 293–305. <http://www.icir.org/vern/imw-2002/imw2002-papers/144.ps.gz>
- [31] Liu X, Ravindran K, Loguinov D. A queueing-theoretic foundation of available bandwidth estimation: Single-Hop analysis. *IEEE/ACM Trans. on Networking*, 2007,15(4):918–931.
- [32] MAWI. <http://tracer.csl.sony.co.jp/mawi/>
- [33] Tcpreplay. <http://tcpreplay.synfin.net/trac/wiki/>

附中文参考文献:

- [23] 刘敏,李忠诚,过晓冰,邓辉.端到端的可用带宽测量方法. *软件学报*,2006,17(1):108–116. <http://www.jos.org.cn/1000-9825/17/108.htm>
- [24] 刘星成,何莉,余顺争.网络可用带宽的高精度测量算法. *电子学报*,2007,35(1):68–72.



何莉(1982—),男,江西萍乡人,博士生,主要研究领域为网络测量,网络安全,服务质量.



余顺争(1958—),男,博士,教授,博士生导师,主要研究领域为 Internet 流量测量、分析、建模,统计异常检测,隐马尔可夫模型算法,无线网络.