软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2017,28(1):105−134 [doi: 10.13328/j.cnki.jos.005127]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

IP 网络性能测量研究现状和进展*

胡治国[1], 田春岐[2], 杜 亮[1], 关晓蔷[1], 曹 峰[1]

[1](山西大学 计算机与信息技术学院,山西 太原 030006)

[2](同济大学 电子与信息工程学院 计算机科学与技术系,上海 201804)

通讯作者: 胡治国, E-mail: huzhiguotj@sxu.edu.cn

摘 要: 网络性能测量是网络测量领域的核心分支,是指遵照一定的方法和技术,利用软、硬件工具来测试、验证及表征网络性能指标的一系列活动总和,是量化网络性能指标、理解和认识网络行为最基本和最有效的手段,在网络建模、网络安全、网络管理和优化等诸多领域均有广泛应用,是计算机网络领域持续的研究热点之一.介绍了该领域的研究现状与进展,重点讨论了带宽、丢包和时延测量等方面的代表性算法,从算法的基本思想、关键技术、实现机理入手,剖析了突发性背景流的时间不确性和多跳网络路径下的空间不确定性对带宽测量的影响、丢包测量中应用流丢包与探测流丢包的区别与联系、时延测量中时钟偏差与时钟频差的相互作用关系等问题,并在此基础上对网络性能测量面临的挑战、发展趋势和进一步研究的方向进行了讨论.

关键词: 网络性能;带宽;瓶颈定位;丢包;时延

中图法分类号: TP393

中文引用格式: 胡治国,田春岐,杜亮,关晓蔷,曹峰.IP 网络性能测量研究现状和进展.软件学报,2017,28(1):105−134. http://www.jos.org.cn/1000-9825/5127.htm

英文引用格式: Hu ZG, Tian CQ, Du L, Guan XQ, Cao F. Current research and future perspective on IP network performance measurement. Ruan Jian Xue Bao/Journal of Software, 2017,28(1):105−134 (in Chinese). http://www.jos.org.cn/1000-9825/5127.htm

Current Research and Future Perspective on IP Network Performance Measurement

HU Zhi-Guo[1], TIAN Chun-Qi[2], DU Liang[1], GUAN Xiao-Qiang[1], CAO Feng[1]

[1](School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)

[2](Department of Computer Science and Technology, School of Electronics and Information, Tongji University, Shanghai 201804, China)

Abstract: Network performance measurement, which takes advantage of specific methods and techniques to quantify network performance, is the core branch of the network measurement research. Network performance measurement and analysis provides the most effective and fundamental way of quantifying network performance and characterizing network behavior. Network performance measurement has great significance in network modeling, network security, network management and network optimization. Drawing on the latest research progress, this paper presents the principles, characteristics and implementation mechanisms of the representative algorithms for bandwidth measurement, packet loss measurement and delay measurement. The main discussions include temporal uncertainty and spatial uncertainty during bandwidth measurement, the distinctness of application packet loss and probe packet loss, and the relationship between the clock offset and clock skew. Lastly, the paper discusses some challenges and directions in the field network performance measurement study.

Key words: network performance; bandwidth; bottleneck location; packet loss; packet delay

* 基金项目: 国家自然科学基金(61502289, 41401521); 山西省青年科技研究基金(2015021101)

Foundation item: National Natural Science Foundation of China (61502289, 41401521); Natural Science Foundation for Young Scientists of Shanxi Province (2015021101)

收稿时间: 2016-03-03; 修改时间: 2016-06-14; 采用时间: 2016-07-28; jos 在线出版时间: 2016-10-11

CNKI 网络优先出版: 2016-10-12 16:26:51, http://www.cnki.net/kcms/detail/11.2560.TP.20161012.1626.018.html

© 中国科学院软件研究所 http://www.jos.org.cn

以 Internet 为代表的 IP 网络是复杂的巨系统,是人类社会信息化的标志之一,其行为影响着我们每个人的工作和生活.对 Internet 的运行管理和管治,无论从社会、商业和技术的角度来看都愈益重要和迫切.网络测量作为监控、理解和认识网络行为,进而优化和重新规划网络结构以及改善网络服务质量的重要手段,受到了工业界和学术界的广泛重视[1,2].概括来讲,针对 Internet 的网络测量研究,其意义主要体现在以下几个方面:(1) 为网络行为学的研究提供重要的理论依据和准确的验证平台,是对理论模型进行验证与修正的重要基准;(2) 通过网络测量可判断和评估当前网络对应用的支持程度;(3) 可实时检测网络拥塞状况、定位网络性能瓶颈,从而及时了解网络运行状况;(4) 利用网络测量方法和对网络拓扑变化的分析,可对网络异常进行提前告警和建立有效的网络安全防范机制,保证网络安全、稳定地运行[3,4].

1996 年,美国应用网络研究国家实验室(NLANR)组织召开了互联网统计与指标分析(ISMA)研讨会,这标志着大规模、系统化网络测量研究的开始.一批在网络测量领域有重大影响的项目应运而生,如 NIMI[5],Surveyor[6] 和 CAIDA[7]等.我国在该领域的研究起步相对较晚,研究成果主要来自湖南大学、国防科学技术大学、电子科技大学、中国科学院等单位.近年来,我国科研人员陆续在国际顶尖学术会议和期刊 MOBICOM[8],SIGCOMM[9],INFOCOM[10],TON[11],JSAC[12],TMC[13]上发表了许多与网络测量相关的研究成果,如文献[14–19]等,标志着我国在网络测量领域取得了长足的进步.

网络测量的分类标准有很多种:按测量方式不同,可分为主动测量和被动测量;按测量环境不同,可分为单点测量和多点测量;按测量的对象不同,可分为网络拓扑测量、网络性能测量和网络流量测量;按测量点所在层次不同,可分为网络层测量和应用层测量;按测量者是否知情,可分为协作式测量和非协作式测量;按测量所采用的协议,可分为基于 BGP 协议、基于 TCP/IP 协议和基于 SNMP 协议的测量,等等.其中,对网络性能测量的研究最为集中,影响也最为广泛.虽然不同组织和文献对网络性能参数的定义不尽相同,但绝大多数的研究都将带宽、丢包、时延和抖动作为评价网络性能的基本参数指标,并据此分析网络的连通性、可靠性、稳定性和安全性.上述参数中,抖动是时延值变化情况的体现,只需通过对时延测量结果的分析就能得到对应的抖动值.因此,网络性能测量又可具体分为带宽测量、丢包测量和时延测量.

针对网络性能测量,国内已有一些综述性文章,如文献[20–22]等,但主要关注于带宽测量算法的介绍,对影响算法性能的关键因素,如突发性背景流下的时间不确性和多跳网络路径下的空间不确性对带宽测量的影响缺乏认识;对丢包测量和时延测量算法目前还鲜有文献进行介绍和剖析.基于上述考虑,本文以带宽、时延、丢包测量中的代表性算法为主要研究对象,以此梳理各测量技术涉及到的基本概念、算法思想、实现机理和存在的问题,并对未来网络性能测量的研究方向进行了讨论.

1 带宽测量算法

1.1 基本概念

定义 1(背景流量(cross traffic)). 这是指网络路径上已经存在的流量.若背景流量在任意时刻的传输速率保持不变,则称背景流为恒定背景流或流体模型,其他类型背景流则称为突发背景流或非流体模型.

定义 2(链路带宽(link bandwidth)). 链路带宽又称为链路容量(link capacity),是指在无背景流量条件下,链路所能提供的最大传输速率.

定义 3(路径带宽(path bandwidth)). 路径带宽又称为路径容量(path capacity),是指在无背景流量条件下,网络路径所能提供的最大传输速率.

定义 4(可用带宽(available bandwidth)). 这是指在不影响背景流传输速率的情况下,链路能为其他应用提供的最大传输速率.文献[23]对流体模型和非流体模型下的可用带宽分别进行了定义,即,流体模型下链路的可用带宽为 $avbw=Capacity-CrossTraffic_{\Delta t}$,$Capacity$ 为链路容量,$CrossTraffic_{\Delta t}$ 为 Δt 时间内通过该链路背景流的平均速率;非流体模型下链路的可用带宽: $avbw_{\Delta t}=\frac{LC_{\Delta t}-CT_{\Delta t}}{\Delta t}$,$LC_{\Delta t}$ 为在时间段 Δt 内能通过链路 L 的总比特数,$CT_{\Delta t}$ 为在时间段 Δt 内流过链路 L 的比特数.

© 中国科学院软件研究所 http://www.jos.org.cn

定义 5(窄链路(narrow link)). 这是指网络路径上容量最小的链路.

定义 6(紧链路(tight link)). 紧链路又称为瓶颈链路,是指网络路径上可用带宽最小的链路.

在网络路径上,窄链路和紧链路可能为同一链路,也可能是不同的链路.

1.2 链路/路径带宽测量

根据探测包发送方式的不同,链路/路径带宽测量算法分为可变包长(variable packet size,简称 VPS)技术和测量包对(packet pair,简称 PP)技术两类.其中,基于可变包长技术的算法主要有 pathchar[24],pchar[25],clink[26],基于测量包对技术的算法主要有 Bprobe[27],pathrate[28],CapProbe[29],nettimer[30]和 SigMon[31]等.

1.2.1 可变包长度技术

可变包长测量技术(VPS)主要测量网络路径上的单跳带宽.算法思想如下:

- VPS 方法向测量路径中发送 TTL(time to live)受限制的探测包,通过 TTL 值,使得探测报文在指定路由器上发生超时,并向源端返回一个 ICMP TTL 超时报文,则源端可通过收到的 ICMP 报文来计算到达指定链路的往返时延 RTT(如图 1 所示).

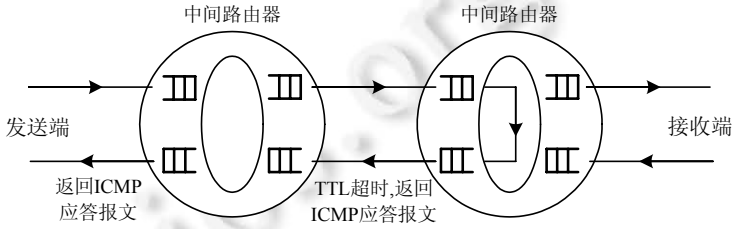


Fig.1 Network model[24]

图 1 网络模型[24]

- 从发送端到路径每跳的最小 RTT 中不包含探测包在路由器中的排队时间,其与探测包大小成正比,与链路带宽成反比;通过线性回归技术逐跳地测量路径上每一跳链路的容量(如式(3)和图 2 所示).

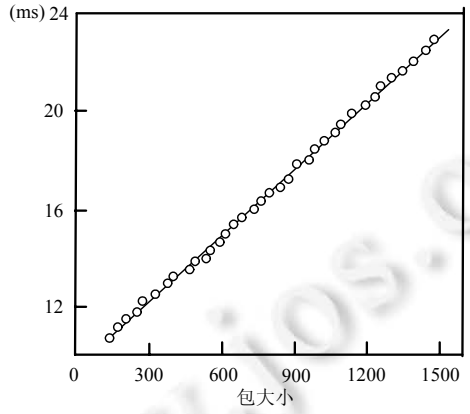


Fig.2 Shortest observed RTT vs. packet size[24]

图 2 最小时延与包大小相互关系[24]

具体过程如下:

设从路径源端发送长度为 L 的探测报文,TTL 为 n,则往返时延为

© 中国科学院软件研究所 http://www.jos.org.cn

$$RTT_n=\sum_{i=1}^{n}\left(\frac{L}{C_i}+\frac{d_i}{v_i}+p_i+q_i\right)+\sum_{i=1}^{n}\left(\frac{L'}{C_i'}+\frac{d_i'}{v_i'}+p_i'+q_i'\right) \tag{1}$$

其中,L 为探测报文的长度,C_i 为第 i 跳链路的容量,d_i 为第 i 跳链路的长度,v_i 为信号传播速度,p_i 为探测报文在第 i 跳路由器的处理时延,q_i 为探测报文在第 i 跳路由器的排队时延,L'为返回应答报文的长度,C_i',d_i',v_i',p_i',q_i' 为返回路径上对应的性能参数.

VPS 方法认为:(1) 通过多次测量,可获得排队时延为 0 的探测包,即,多次测量可以得到最小时延;(2) 相同路径中,其传播时延相同;(3) 路由器处理时延相对固定;(4) 返回应答报文的长度固定.

基于上述条件,公式(1)可化简为

$$\min(RTT_n)=\sum_{i=1}^{n}\left(\frac{L}{C_i}+\frac{d_i}{v_i}+p_i\right)+\sum_{i=1}^{n}\left(\frac{L'}{C_i'}+\frac{d_i'}{v_i'}+p_i'\right)=\sum_{i=1}^{n}\left(\frac{d_i}{v_i}+p_i+\frac{L'}{C_i'}+\frac{d_i'}{v_i'}+p_i'\right)+\sum_{i=1}^{n}\left(\frac{L}{C_i}\right)=a_n+\beta_n L \tag{2}$$

其中,$\beta_n=\sum_{i=1}^{n}\left(\frac{1}{C_i}\right)$.由此可见,$\min(RTT_n)$是由与报文长度 L 有关的以及与报文长度无关的两部分构成.对于每个确定的 n,不同大小的 L 值对应有不同大小的 $\min(RTT_n)$.$\min(RTT_n)$是 L 的线性函数,β_n 为函数的斜率.通过向网络发送长度不同的探测报文,在发送端测量每个探测报文的最小往返时延,能获得一系列的样本点(L, $\min(RTT_n)$).通过这些样本点,可以画出一条直线,斜率为β_n,如图 2 所示.同理,也可得到β_{n-1},然后通过公式(3)即可得到链路 L_n 的容量 C_n:

$$C_n=\frac{1}{\beta_n-\beta_{n-1}} \tag{3}$$

Pathchar 算法是 VPS 方法中最具代表性的算法,pchar 和 clink 的算法思想均衍生于 Pathchar.

VPS 算法的主要不足表现在:VPS 方法需向被测链路发送多组探测包,从中获取最小时延,并提高测量精度,但不可避免地会增加链路负载.随着链路跳数的增加,测量所需探测包的数量也会急剧增加,测量过程中如果链路出现拥塞而产生排队,测量结果将会出现严重的失真.此外,如果网络路径中存在第 2 层交换设备,则 VPS 算法不适用于此类路径的带宽测量.这是因为两层交换设备不修改探测报文的 TTL 值,不会向源端响应 TTL 超时报文,且交换设备会给探测包增加一个与探测包长度成正比的传输时延,导致实际测量到的时延与 VPS 技术所依赖的时延模型不相符而产生较大的测量误差.

1.2.2 测量包对技术

测量包对技术(PP)主要用于测量路径带宽.算法基本思想是:源端向目的端背靠背地发送探测包对,假设背靠背的测量包对仅在瓶颈链路处发生排队,即,在瓶颈链路之前和之后的链路中都无排队现象出现;经过瓶颈链路后,测量包对间的间隔时间与包长度成正比,与路径容量成反比,以此来计算路径带宽.

由于包对在源端是以背靠背的方式发送,故包对离开源端时的散布时间为 L/C_0;当包对经过路径第 i 个链路时,设该跳链路容量为 C_i,并且链路负载为 0,则包对在该跳链路之前的散布时间Δ_{in}和离开该跳链路之后的散布时间Δ_{out}满足公式(4):

$$D_{out}=\max\{D_{in},L/C_i\} \tag{4}$$

包对进入/离开链路 i 的散布时间关系如图 3 所示.

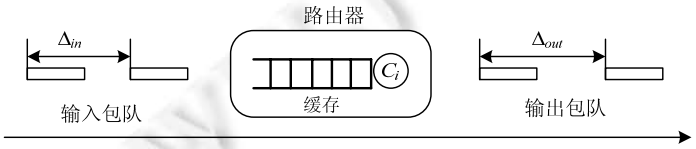


Fig.3 Dispersion time at link i (in/out)[28]

图 3 包对进入/离开链路 i 的散布时间关系[28]

若假设包对经过路径中每跳链路时链路都不存在背景流量,则包对在目的端的散布时间为

© 中国科学院软件研究所 http://www.jos.org.cn

$$\Delta_R = \max_{i=0,\ldots,H}\{L/C_i\} = \frac{L}{\min\limits_{i=0,\ldots,H}(C_i)} = \frac{L}{C} \tag{5}$$

其中,L 为探测包长度,C 为路径带宽.

在目的端可以记录每个探测包的到达时间,据此可以计算包对在目的端的散布时间Δ_R,因此,可以根据公式(6)来测量路径容量.

$$C=L/\Delta_R \tag{6}$$

包对技术算法的主要不足为:包对技术受背景流量的影响非常严重,如果背景流量干扰使得第 1 个测量包出现延迟,则会导致两个测量包之间的间隔时间被压缩,那么测量所得链路带宽值就会偏高;如果干扰流量出现在第 1 个测量包和第 2 个测量包之间,那么将会出现排队的情况,则两个测量包之间的间隔将会被拉大,测量所得带宽值就会偏小.

实际测量时,背景流量不可避免地会影响到测量包对.对于一个有 n 个测量样本的集合,采用何种方式或规则统计和过滤测量样本,是包对技术的关键.Bprobe 算法发送具有不同包长度的探测包对,对测量样本进行并集和交集的过滤处理来得出最终的测量结果.Nettimer 算法假设路径容量是测量样本分布的峰值,提出核密度估计函数的样本过滤方法.Pathrate 算法首先利用数据包对进行采样,若样本呈多峰分布,则使用包列过滤多峰分布的影响,进而计算路径的渐进分离速率 ADR(asymptotic dispersion rate),但 ADR 并非路径带宽.CapProbe 算法提出利用双包时延和(两个包的单向时延之和)来过滤测量样本,逐一记录所有的双包时延和以及对应的双包间隔 T 值,在估测带宽值时,找出时延和最小的探测包样本,然后用该样本相应的 T 值来估计带宽.CapProbe 算法综合分析了包对时延、包对间隔等因素对测量精度的影响,有效剔除了失真样本,文献[29]的实验结果表明,CapProbe 算法的测量精度和收敛性能均优于 Pathrate 算法.

Nettimer 算法综合利用了变包长测量技术和包对技术,提出了 PT(packet tailgating)模型.其基本思想是:向路径中注入两个连续的探测包,首先发送尽量大的探测包(以该包在网络传输过程中不被分片为约束条件),该包带有一个 TTL 值(使其到指定链路时的 TTL 变为 0),然后这个包后面紧跟着发送一个小的包.由于小包的传输延迟远远小于第 1 个大包的传输延迟,这就导致小包(称为 tailgater)在大包(称为 tailgated)后面阻塞排队,而由于大包在指定链路处被丢弃,因此小包在此后的链接中不会因阻塞而发生排队.最后从 TTL=1 开始,逐跳取小包的最小时延值,参照 pathchar 方法计算带宽算法,测量各链路的带宽.SigMon 算法在测量路径容量时的算法思想与图 3 所示一致,其算法的主要特点是在测量路径瓶颈容量的同时,实现了对可用带宽的测量.

文献[32]对 pathchar,clink 和 pathrate 算法从测量精度和测量负载等方面进行了对比分析,指出:pathrate 测量精度高于其他算法,但测量负载较高,测量时间较长.文献[33]借鉴包对算法思想(见式(6)),利用 TCP 流的 ACK 包进行路径容量测量,不引入额外测量负载,因此适合长时间地对网络性能进行监测.但由于包对算法本身的缺点,导致该算法在突发背景流下测量误差较大.

1.3 可用带宽测量

可用带宽测量是网络测量研究领域中的重点,拥有数量众多的算法.绝大多数可用带宽测量方法可划分为探测间隔模型(probe gap model,简称 PGM)和探测速率模型(probe rate model,简称 PRM)[34−37].PGM 的代表性算法有 Jitterpath[36],Spruce[37],CProbe[38],IGI[39],Abing[40],AProbing[41];PRM 具有代表性的算法有 SLDRT[23],PTR[39],pathload[42],pathChirp[43],TOPP[44],Yaz[45],等等.

1.3.1 包间隔模型

PGM 的算法思想是:自源端向目的端发送测量包,当测量包的速率大于可用带宽时,在瓶颈链路上的探测包就会发生排队现象,测量包之间的间隔就会发生变化.通过研究这种变化,就能得到链路可用带宽值(如式(7)和图 4 所示).PGM 算法通常假设窄链路和紧链路为同一链路,且窄链路容量已知.

PGM 算法的基本原理如下:

© 中国科学院软件研究所 http://www.jos.org.cn

$$A = C \times \left(1 - \frac{\Delta_{out} - \Delta_{in}}{\Delta_{in}}\right) \tag{7}$$

其中,A 为可用带宽,C 为链路带宽,Δ_{out} 为输出时间间隔,Δ_{in} 为输入时间间隔.

CProbe 算法是最早的可用带宽测量方法,利用了 ICMP 协议的请求-应答机制从发送端向接收端发送多组探测包,并触发节点的 ICMP 处理机制,使得路径上的节点向发送端发送 TE(ICMP time-exceeded)和 DU(ICMP destination-unreachable)包;CProbe 算法接收到 TE 和 DU 包后,根据接收时间推算探测包在每段链路上传输速率的变化情况,进而估算可用带宽.然而,Dovrolis 等人在文献[46]中指出:CProbe 所度量的并不是可用带宽,而是处于可用带宽和带宽容量之间的非对称分散速率 ADR(asymptotic dispersion rate).

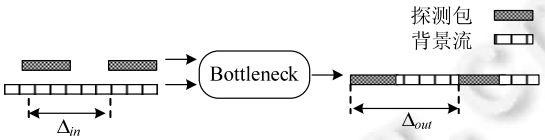


Fig.4 PGM for estimating available bandwidth[37]

图 4 PGM 算法测量可用带宽原理[37]

Spruce 算法根据公式(7)计算可用带宽,发送一串包对取其测量平均值作为最终测量结果.Spruce 算法把测量包对的发送间隔设为瓶颈链路上的传输时延,确保测量包对内部之间的排队队列不为空.两对包对之间的时间间隔符合 Possion 分布,并且令 Possion 分布的期望远大于Δ_{in},从而保证了探测流不会引起瓶颈链路的过度拥塞而导致丢包现象的出现,使 Spruce 具有较轻的测量负载.与 Spruce 算法相比,IGI(initial gap increasing)算法的测量稳定性更好,因此针对 IGI 算法的研究和改进也较多,如文献[47,48].IGI 算法将探测包划分为分离排队区域(disjoint queueing region,简称 DQR)和联合排队区域(joint queuing region,简称 JQR).DQR 表示网络队列正处于缩减阶段,探测包对之间可能没有填充背景流,DQR 区域探测包输出间隔不受背景流量的影响;JQR 表示网络队列长度正处于增长阶段,探测包对之间被背景数据包填充,JQR 区域探测包输出间隔和背景流量相关(如图 5 所示).

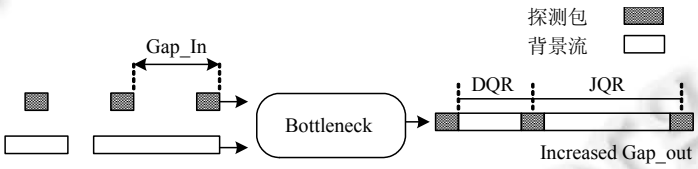


Fig.5 JQR vs. DQR[39]

图 5 分离排队区域与联合排队区域[39]

IGI 算法中,网络负载平均速率 B_C 计算公式为

$$\frac{B_O \sum_{i=1}^{M} (g_i^+ - g_B)}{\sum_{i=1}^{M} g_i^+ + \sum_{i=1}^{K} g_i^= + \sum_{i=1}^{N} g_i^-} \tag{8}$$

其中,$g_i^+, g_i^=, g_i^-$ 分别表示增加、减小和不变的包对间隔,g_B 表示探测分组在瓶颈链路上的传输时间,$B_O \sum_{i=1}^{M} (g_i^+ - g_B)$ 代表探测期间经过路由器的所有背景流量;$\sum_{i=1}^{M} g_i^+ + \sum_{i=1}^{K} g_i^= + \sum_{i=1}^{N} g_i^-$ 代表探测过程中耗费的时间,M,N,K 则对应于探测流中间隔增大、减小和不变的包对个数;B_O 为瓶颈带宽的大小,B_O–B_C 即为可用带宽.

与传统 PGM 算法不同,文献[36]提出一种基于 QDPM(queuing delay propagation model)的可用带宽测量方法,称为 JitterPath.JitterPath 算法发送一系列包对序列来捕捉背景流,并通过计算捕获流量比 CTR(captured traffic ratio)来反映路径上的拥塞状况.当 CTR 大于 1 时,表明测量时网络正处于拥塞状态,这说明探测速率大于可用带宽,反之亦然.根据以上信息,JitterPath 算法采取二分搜索策略调整探测包对序列的发送速率,不断逼近可

© 中国科学院软件研究所 http://www.jos.org.cn

用带宽的真实值.然而,JitterPath 算法需假设可用带宽在测量过程中保持不变,在突发性较强的背景流下,这种假设通常是不成立的.

PGM 算法单次测量时间较短,注入网络的流量较小,因此通常不会对网络造成较大负载.其主要不足为:PGM 通常假设窄链路和紧链路为同一链路,且窄链路容量必须提前获知,但在实际网络中,由于交叉背景流的存在, PGM 模型假设窄链路与紧链路为同一链路的前提条件不能总是成立;PGM 算法假设路径上的背景流是基于流体模型的,这种模型假设背景流速率变化比较缓慢而且在测量时间内速率是恒定的,所以当路径上有突发性背景流时,测量精度无法保障.文献[49]的实验结果也表明:PGM 算法在多瓶颈链路下的测量结果会出现较大偏差,且会低估可用带宽.

1.3.2 包速率模型

PRM 算法思想又称为自拥塞原则,其核心思想是:当所发送的探测流速率高于实际待测链路可用带宽时,则该探测流的单向时延将表现出一种递增趋势;相反地,如果发送速率小于可用带宽,则探测流的单向时延会呈现一种相对平稳的趋势(如图 6 所示),通过对时延变化趋势的判断,寻找发送速率和到达速率开始匹配的转折点,将对应探测流的平均速率作为路径可用带宽的测量值.

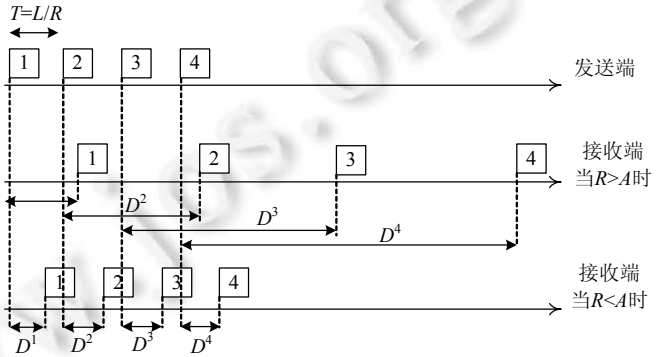


Fig.6 Periodic stream illustration of basic idea[42]

图 6 自负载周期流的基本思想[42]

PRM 算法中,以自负载周期流(self-loading periodic streams,简称 SLoPS)方法最具代表性,其形式化表示为:设一个探测流包含 K 个数据包,每个包大小为 L 个字节,发送端以速率 R_0 在任意时刻开始发送,则包的发送周期为 $T=L/R_0$.包 k 从发送端到接收端的相对单向时延为 $D^k=\sum_{i=1}^{H}(L/C_i+q_i^k/C_i)=\sum_{i=1}^{H}(L/C_i+d_i^k)$,$q_i^k$ 为第 k 个包到达链路 i 时的队列长度(不包括第 k 个包),L/C_i 和 d_i^k 分别为探测包在第 i 条链路上的发送时延和排队时延.前后相邻的两个探测包 k 和 k+1 之间的单向时延差为 $\Delta D^k=\Delta D^{k+1}-D^k=\sum_{i=1}^{H}\Delta q_i^k/C_i=\sum_{i=1}^{H}\Delta d_i^k$, 即:

$$\Delta q_i^k=q_i^{k+1}-q_i^k,\Delta d_i^k=\Delta q_i^k/C_i.$$

可以看出:相邻两个探测包的单向时延差值只和探测包在路由器上的排队时延相关,而排队时延是由路径的拥塞状况所决定的,所以探测包的排队时延反映了在探测期间内网络路径的拥塞状况,即:当探测流的单向时延呈现上升趋势时,可以推断出探测流速率 R_0 大于可用带宽;当探测流中单向时延不呈现出上升趋势时,R_0 小于等于可用带宽.也就是说,通过探测流单向时延的变化情况,可分析出探测流速率与可用带宽的匹配程度,进而使之逐渐逼近可用带宽.

Pathload 和 pathChirp 是 PRM 模型中最具代表性的算法.文献[50]对 10 余种常用的可用带宽测量算法进行对比后发现:pathload 算法的测量精度最高,pathChirp 算法对网络造成的拥塞时间最短.Pathload 通过 S_{PCT} (pairwise comparison test)和 S_{PDT}(pairwise difference test)参数来判定探测流的单向时延趋势.S_{PCT} 表示探测流中

© 中国科学院软件研究所 http://www.jos.org.cn

单向时延连续上升的探测包占所有探测包的比例.S_{PDT} 表示探测流单向时延从开始到结束时的变化强度.当 $S_{PCT}>0.55$,$S_{PDT}>0.4$ 时,认为探测流的单向时延具有增长趋势;否则,表明探测流的单向时延没有增长趋势.发送端根据探测流单向时延的变化趋势,按照二分搜索法选择下一次的发送速率.Yaz 算法是针对 Pathload 算法的一种改进,对网络路径是否处于拥塞状态的判定条件进行了调整,与 pathload 算法相比,测量精度基本相同,测量负载有所降低.

pathChirp 算法以指数增长的方式在单条包列内快速提高发送速率(包列构造如图 7 所示,在关于 pathChirp 的文献中,将单条测量包列称为 chirp),用单条包列即可测量出可用带宽.算法假设在测量开始时 chirp 的发送速率低于路径的可用带宽,因此其时延不会有明显的变化;而随着相邻测量包之间的间隔距离的降低,chirp 的发送速率逐渐接近路径的可用带宽,并最终超过可用带宽,这就会造成 chirp 中探测分组的时延出现单调增长的情况.pathChirp 通过分析 chirp 分组的排队时延图形(也称排队时延偏移)来估计端到端路径的可用带宽.一个典型的排队时延偏移包含一块排队时延增长的区域,这是由于探测分组遇到了一个暂时阻塞的队列.排队时延偏移中也可能包含一些排队时延降低的区域,这是因为探测分组遇到了正在清空的队列(如图 8 所示).利用对时延图形的分析,pathChirp 计算每个探测报文的瞬间可用带宽,再通过一个加权公式来计算 chirp 测得的可用带宽,记为 D^m($D^m=\sum_{j=1}^{N-1}E_j^m\Delta_j\Big/\sum_{j=1}^{N-1}\Delta_j$, Δ_j 为包间隔, E_j^m 为报文 i 的瞬间可用带宽),然后,使用一个滑动窗口对数个 chirp 测得的可用带宽加以平均,便得到测量结果.

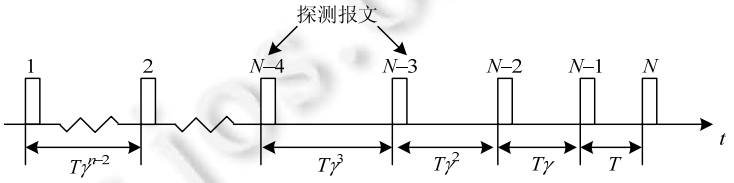


Fig.7 PathChirp probe train[43]

图 7 PathChirp 探测包列结构[43]

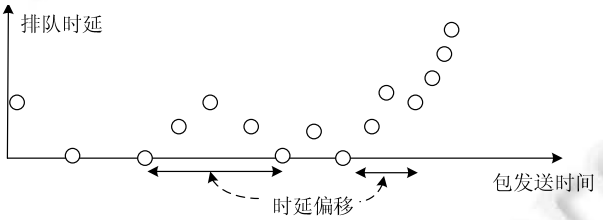


Fig.8 A typical chirp queuing delay signature[43]

图 8 chirp 之中典型的排队时延特性[43]

TOPP 算法将探测流速率划分为 N 个等级($N=\lceil(R_{max}-R_{min})/\Delta_R\rceil$,$R_{min}$ 为最低发送速率,R_{max} 为最高发送速率,Δ_R 为划分间隔),从低到高逐渐增加探测速率.在每个速率等级上,发送 n 个大小相同且时间间隔相同的探测包.当最大注入流量速率与接收端接收到的流量速率近似相等时,则认为此时注入流量的速率即为被测路径的可用带宽.SLDRT 算法与 pathChirp 算法的测量包列构成相反,SLDRT 算法首先背靠背地发送 d 个负载包(loading packet),随后以指数形式逐步降低包列的输入速率,当包列的输入速率等于输出速率时测量结束,以此时测量包列的输入速率作为测量结果输出.由于背靠背的测量包在突发背景流下引入的测量噪声较小,所以 SLDRT 测量算法精度较高,但对网络造成的测量负载高于 pathChirp.

PRM 算法的不足之处是:尽管 PRM 算法在测量时不需要窄链路和紧链路为同一链路这一假设前提,但其不可避免地会引入较多的测量负载,从而造成路径的短暂拥塞.与 PGM 算法一样,大多数 PRM 算法也假设背景流为流体模型,当路径上有突发性背景流时,测量精度仍无法保证.

近年来,完全创新的可用带宽测量方法已较少出现,即使最新的研究成果也多是基于已有算法思想的改进.

© 中国科学院软件研究所 http://www.jos.org.cn

如:文献[41]提出了 AProbing 算法,对 TCP 协议中确认包(ACK packet)进行了重新构造,利用两个连续 ACK 包获得包间隔的变化信息,然后依据 PGM 算法思想对可用带宽进行测量,与典型 PGM 算法对比,精度相同,测量负载减小;文献[51]提出了 GNAPP 算法,利用 ICMP 的应答信息来确定路径跳数和测量可用带宽,测量包列由多个包对组成,包对之间的间隔以线性方式增长,本质上也是利用 PRM 算法思想来测量可用带宽;文献[52]提出了以最小往返时延(RTT)值出现的频率来估计可用带宽,但最小 RTT 的判断阈值在该文献中缺乏说明;文献[31]提出了 SigMon 算法,探测包列由包间隔等比增加的 30 个包对组成,利用包间隔变化幅度和拐点变化情况来分析网络瓶颈带宽和可用带宽,其算法思想可看作是路径带宽的测量包对技术和可用带宽测量的 PGM 模型的综合体.

1.4 紧链路定位

紧链路定位是与可用带宽测量紧密相关的一个研究领域.紧链路定位方法主要有 BFind[53],DRPS[54],STAB[55],pathLoche[56]和 Pathneck[57]等.

Pathneck 是紧链路定位最具代表性的算法.Pathneck 采用递归探测包列 RPT(recursive packets train)定位紧链路,RPT 由负荷分组(load packet)和测量分组(measurement packet)构成(如图 9 所示),RPT 中间的 60 个数据包是负荷分组,每分组 500 字节,RPT 两侧分别是 30 个测量分组,测量分组数 TTL 值可以根据实际路由跳数加以增减.负荷分组能够由发送端到达目的端,用来产生一个长度可测的数据包列.排列在负荷分组列前后的是测量分组,每到达一跳路由器,处于 RPT 首尾的两个测量分组将因 TTL 超时而被丢弃,同时,路由器响应两个对应的 ICMP 分组,这两个应答分组返回源端时的间距被认为是测量分组到达该跳路由的时间间距,反映了 RPT 到达该跳路由时的长度.发送一列 RPT,筛选出 RPT 长度达到最大的转折点链路作为候选紧链路.多次发送 RPT,最后根据置信度统计结果确定紧链路位置.

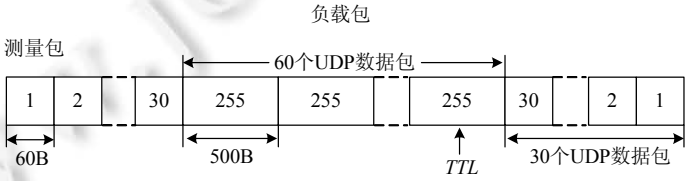


Fig.9 Recursive packet train (RPT)[57]

图 9 递归探测包列(RPT)[57]

尽管 Pathneck 算法无需知道子路径或整条路径的可用带宽便可得到较为准确的定位结果,但与 PGM 模型一样,其算法成立需基于窄链路与紧链路为同一链路的假设前提.此外,Pathneck 测量包基于 ICMP 协议,若路径上的节点不开启 ICMP 功能,则算法无法定位紧链路[58].

BFind 算法通过向网络中发送速率逐渐增大的 UDP 探测包列,进而逐步导致网络产生短时间的拥塞,同时用 traceroute 测量每条链路的往返时延(RTT),若测得的各条链路上的 RTT 值没有增大,则进一步加大探测序列,直至某条链路的 RTT 出现增长趋势,即把此链路确定为紧链路.BFind 算法在定位过程中发送大量的探测包,入侵度非常高,为了避免测量对网络造成较大的影响,BFind 设置能够探测的最大紧链路带宽为 50Mbit/s.DRPS 算法利用自负载周期流特性(SLoPS)与数据包挡板(tailgating)原理[59]相结合,提出一种基于双速率包列的紧链路定位方法,探测包首先以较高的速率通过待测路径,若发生拥塞,则将探测速率调节到一个较低水平,然后通过二分查找算法不断调整探测包的速率,在接收端通过探测包的时延变化和数据包中的 TTL 值来定位紧链路.DRPS 算法的测量负载也较高.STAB 算法利用 pathChirp 算法估计子路径可用带宽,将 pathChirp 探测包改成尾随探测包对,即,每一个探测分组后面紧跟着一个背靠背的探测小分组,此包对中,前导数据包长度较大但时间生存周期较小(TTL 记为 m),而另一个作为尾随包,其长度较小但是有较长的时间生存周期,则在第 m 段路径处大包会被丢弃,而小包则将信息带到接收端.利用这样的尾随包对,STAB 算法可以测量网络路径中任意位置的可用带宽,并判断紧链路的位置.PathLoche 算法借鉴了 STAB 算法思想,其测量子路径可用带宽的算法来源于 SLDRT.STAB 和 PathLoche 算法的测量精度由其采用的可用带宽测量算法所决定,由于它们均要发送多个包列

© 中国科学院软件研究所 http://www.jos.org.cn

来测量不同子路径的可用带宽,因而测量负载通常较大.

1.5 突发性背景流对带宽测量的影响

背景流的特性显然对带宽测量有着重大的影响.目前,绝大多数的带宽测量算法需基于网络背景流为流体模型的假设前提.但文献[60]的研究表明:真实网络流量具有自相似性特性,自相似的网络流量在任意时间尺度上均具有突发性的特点.这个特性在近年来的各类研究中已被多次证明.目前,在路径容量测量及可用带宽测量算法中,仅 Jitterpath 算法和 SLDRT 算法对突发性背景流下算法的可靠性进行了理论上的讨论.

本文将自相似网络流量在任意时间尺度上均具有突发性的特点称为突发性背景流的时间不确定性,在流体模型下,网络的窄链路与紧链路相一致,但由于网络上交叉背景流的存在(背景流的起点和终点都不是路径的终端节点)和背景流量的突发性,紧链路和窄链路可能发生背离,紧链路也可能发生位置变化,本文将其称为多跳网络空间不确定性.突发性背景流时间不确定性、多跳网络空间不确定性的存在,为带宽测量带来了极大的挑战.具体说明如下.

1.5.1 时间不确定性对测量的影响

突发性背景流下,网络路径可用带宽具有时间不确定性的特点,即:无法假设在某个时间粒度内,可用带宽的值保持恒定不变.任何可用带宽测量技术一旦无法在单个采样内得出可用带宽的值,就有可能永远不能正确收敛,尤其是使用二分搜索策略的 pathload,JitterPath 和 Yaz 算法等.如图 10 所示(纵轴为可用带宽,横轴为时间).随着时间的推移,可用带宽在一定范围内波动,假设一条探测包列通过路径时,可用带宽处于高位(时间 t_1 至 t_2),探针包列速率(标为红色)低于可用带宽,则测量算法将该速率标记为可用带宽下界.然而之后的可用带宽始终低于该下界,可用带宽的均值也低于此界,导致 pathload 等二分搜索算法得出了错误的可用带宽下界,致使最终测量值高于真实的可用带宽.反之,当探针包列遇到可用带宽的低谷时(时间 t_3 至 t_4),探针包列的速率高于可用带宽,则可用带宽将该速率标记为上界,但之后的可用带宽始终高于该值,可用带宽的均值也高于此界,最终可能导致 pathload 等二分搜索算法得出了错误的可用带宽上界,从而致使测量值低于真实值.

另一方面,时间不确定性导致可用带宽测量工具可能出现概率性采样误差,即,多次采到峰值或谷值,从而导致平均化后的可用带宽偏离真实数值,这种情况如图 11 所示.这种误差在采样次数足够多的情况下可以被逐渐消除,几种常见的工具如 pathChirp,IGI/PTR 等就采用了多次采样取均值的方法.然而这却极大地延长了测量的时间,从而导致工具不能进行及时、准确的测量.

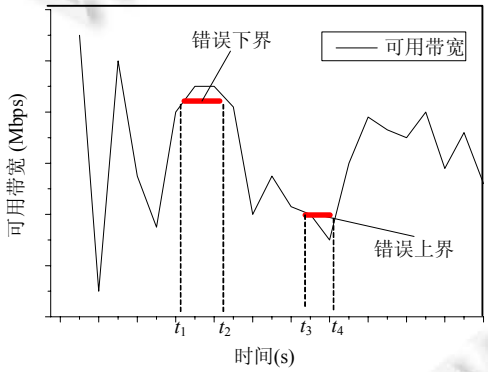


Fig.10 Wrong upper and lower bound caused by temporal uncertainty[23]

图 10 时间不确定性造成的错误上下界[23]

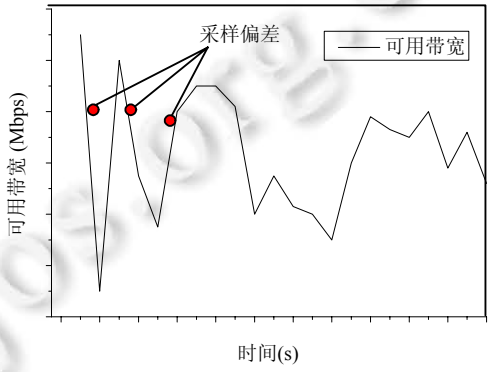


Fig.11 Sampling bias induced by temporal uncertainty[23]

图 11 空间不确定性造成采样误差[23]

1.5.2 空间不确定性对测量的影响

在多跳网络中,路径可用带宽取决于紧链路的可用带宽,一个可能发生的问题是紧链路与窄链路出现背离,即,可用带宽最小的链路并不总是带宽容量最小的链路.如图 12 所示,C_1,C_2 分别代表链路 1(Link 1)和链路 2

© 中国科学院软件研究所 http://www.jos.org.cn

(Link 2)的容量,A_1,$A_{1'}$,A_2,$A_{2'}$,A_3 和 $A_{3'}$分别表示同一时间链路 1 和链路 2 的可用带宽值,正如图 12 所表示的那样,网络中会出现窄链路是 Link 1(C_1<C_2),而紧链路却是 Link 2 的现象($A_{2'}$<A_2).网络流量的突发性使得非紧链路可能在突发时间段内具有比紧链路更小的可用带宽,从而导致 PGM 算法成立的前提条件不满足,进而出现校大的测量误差.

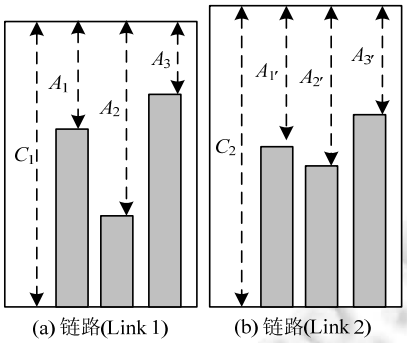


Fig.12 Available bandwidth variation

图 12 可用带宽波动变化

文献[38]首先探讨了紧链路前效应与紧链路后效应,指出,非紧链路的突发性流量对可用带宽测量准确性具有显著影响.文献[60]分别从理论与实践的角度研究了网络路径多跳效应,指出,基于包列的可用带宽测量工具的准确性直接受到包列长度的影响——探测包列越长,测量结果越接近真实的带宽变化曲线.这解释了为什么 IGI 等基于包对的探测技术通常在多跳场景内缺乏健壮性.虽然多数情况下紧链路和窄链路为同一链路,但一个准确和健壮的测量工具不应该基于这种假设,这将使得测量算法不具备对 Internet 和具有突发背景流网络的监测能力.文献[61,62]从理论上研究了背景流对可用带宽测量的影响,指出:突出背景流下测量误差不可能完全消除,但较长的探测包列可减少测量误差和紧链路前效应与紧链路后效应对测量工具准确度的影响.这对设计新的可用带宽测量方法有一定的指导意义.

1.6 算法对比分析

文献[23,50]的研究表明:通常情况下,在众多带宽测量算法中,pathChirp 测量负载最小,pathload 测量精度最高,SLDRT 测量时间最短.因此,以 pathload,pathChirp 和 SLDRT 算法作为分析对象.由于 pathload 测量得到的是可用带宽变化区间[R_{max},R_{min}],为便于统计分析,取(R_{max}+R_{min})/2 为 pathload 测量结果.

设置 NS-2 仿真拓扑环境,流量以 One-hop persistent 方式发送(如图 13 所示),实验结果如图 14 和表 1 所示.

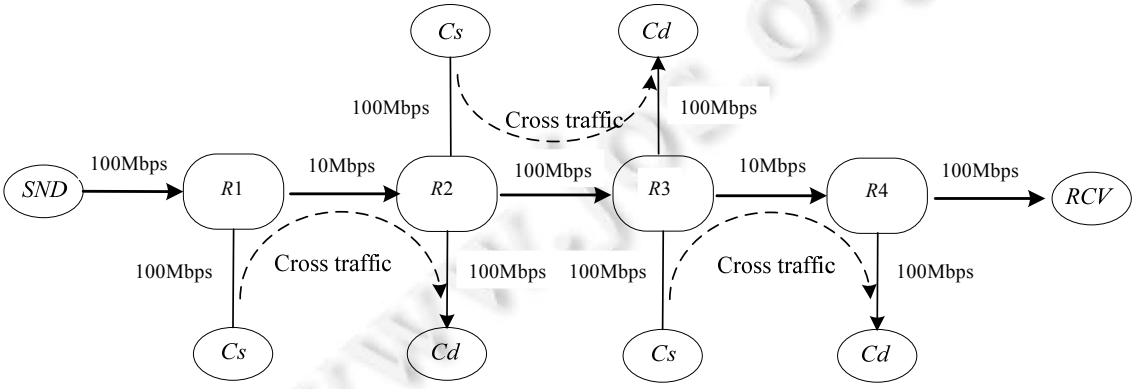


Fig.13 One-Hop persistent model

图 13 One-Hop persistent 背景流

© 中国科学院软件研究所 http://www.jos.org.cn

Table 1 Experimental results

表 1 验证实验结果

方法	测量时间(s)	测量负载(MB)	测量误差(%)
SLDRT	0.088	0.080	7.87
pathChirp	3.69	0.328	12.27
pathload	25.44	2.47	7.26

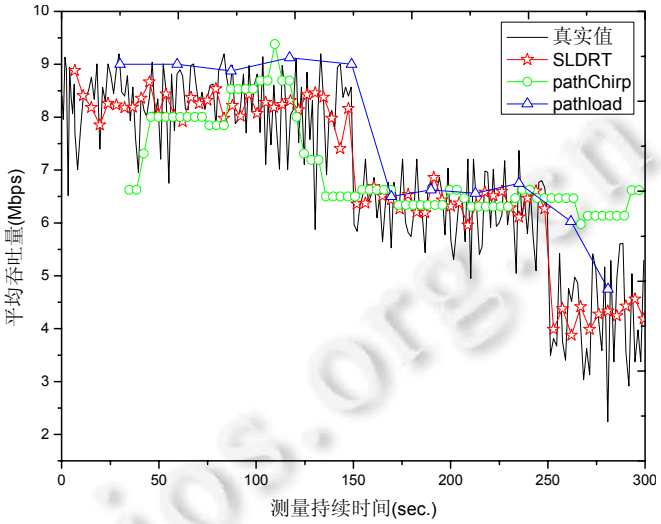


Fig.14 Experimental results

图 14 实验结果对比分析

实验结果表明:各种不同算法反映了带宽变化的总体情况,但测量精度、测量时间、测量负载存在明显差异.Pathload 和 SLDRT 算法测量精度较高,我们分析其原因为:SLDRT 算法以一个较长的包列测量可用带宽,按照文献[61,62]的误差分析理论,长包列的测量误差通常较小;而 pathload 测量值是一个带宽波动区间范围,涵盖面广,将多种可能的测量结果包含其中.而 pathChirp 本质上使用相邻包对的瞬时速率判断带宽值,所以测量精度较差.在测量时间和测量负载方面,SLDRT 测量时仅发送单个测量包列,所以测量时间和测量负载均较小;而 pathChirp 和 pathload 均需通过多个包列才可获得可用带宽(pathChirp 的 Chirp 数默认设置为 7,pathload 算法每个 fleet 中 stream 的数量为 12).此外,pathChirp 和 pathload 测量周期较长,在流量大小发生突变时测量结果就会出现较大的偏差(见实验 150s 和 250s 处),这与第 1.5 节的分析相一致.更加详细、完整的可用带宽算法对比分析结果可参考文献[23,36,45,50].

综上分析,我们认为,利用较长的测量包列来测量带宽是未来的研究方向之一.

2 丢包测量

2.1 基本概念

定义 7(丢包(packet loss)). 源端向目的端发送的分组,若在指定时间内该分组没有到达目的端,则称为分组丢失或丢包.物理线路故障、网络拥塞、设备故障、病毒攻击、路由信息错误等是网络发生丢包的主要原因[63].

定义 8(丢包率(packet loss rate)). 丢失分组与发送分组总数的比值.若 a 表示发送分组总数,b 表示接收到的分组总数,则丢包率为$(a-b)/a$[20].

定义 9(丢包模型(packet loss model)). 网络中分组丢失之间往往存在短暂的相关性,这种相关性可用数学模型来表征,即丢包模型.如,用随机变量 X_i 代表丢包事件,X_i=0 表示数据包丢失,而 X_i=1 表示数据包未丢失.那么第 i 个分组丢失的概率为 $p=[X_i|X_{i-1},X_{i-2},\ldots,X_{i-n}]$,$X_{i-1},X_{i-2},\ldots,X_{i-n}$ 取所有可能的组合情况[20].

© 中国科学院软件研究所 http://www.jos.org.cn

2.2 丢包测量算法

通常,网络中丢包发生次数相对较少且持续的时间很短,因此,测量丢包和确定丢包模型是一件很有挑战性的工作.目前,针对 IP 网络的丢包测量算法主要有 PING[64],ZING[65],Sting[66],BADABING[67]和 PLBU[68]等.PING 发送多个 ICMP 数据包,根据 ICMP 包的应答信息来判断是否发生丢包.显然,应答包的丢失也会直接影响测量精度.此外,如果传输过程中防火墙或路由器启用了 ICMP 过滤功能,则 PING 方法就无法使用.ZING 算法是基于泊松模型 PASTA(Poisson arrivals see time averages)的丢包测量方法,其使用概率方法独立且随机地发送探测数据包,对网络的入侵度较小.Sting 算法利用 TCP 协议的通信过程来测量丢包率,可测量前向路径(forward path)和反向路径(reverse path)上的丢包情况,但其主要针对基于 TCP 协议的应用.尽管 RFC 2330[69]建议以泊松采样作为分组丢失测量的基本方式,但文献[67]的研究指出:泊松采样虽然是一种渐近采样方式,但是,由于丢包属于小概率事件,要准确地测量丢包就意味着需增加测量时间,或者提高探测流的发送速率以及时发现丢包,这将不可避免地增大网络的额外负载.文献[67]利用排队论原理提出了 BADABING 丢包测量方法(包列构造如图 15 所示),其主要思想如下:探测包列由多对背靠背的探测包对组成,探测包对中背靠背包的数量可任意设置,包对的发送间隔固定,但每个包对在测量中是否发送由一个固定概率决定.若单个探测包在接收端未被接收到,则标记为 1,否则,记为 0.当包列内出现连续两个丢包(将其标记为 11)时,则认为测量路径开始丢包,直到收到包的标记为 10 或 00 时,则认为丢包结束,并以此计算丢包的持续时间.探测包列中,包对以给定概率的方式发送,从而较好地平衡了测量准确性和测量入侵性这对矛盾.文献[67]中的实验表明:在相同发送速率下,BADABING 比标准泊松测量包有着更高的测量准确性,且可推测丢包持续时间.

通常,在丢包测量过程中,为了更好地分析数据包丢失情况且不失一般性,认为探测流(probe traffic)的丢包率就是背景流(cross traffic)的丢包率,并反映当前应用流的丢包情况,且假设注入的测量数据包不会引起背景流额外的丢包.即使 BADABING 算法通过设置包列中探测包数量的方法较好地实现了对丢包持续时间的测量,也不可能准确捕获测量过程中所有的丢包情况.因此,其测量结果并不一定与背景流的丢包情况相一致.此外,这种不断注入探测流的测量方法不可避免地对网络造成较大的额外负载.针对上述问题,文献[68]提出了 PLBU 算法.PLBU 算法利用 MPEG 4 或 H.264 视频中用户数据域(user_data field)进行丢包测量,其算法思想是:在视频文件传送之前,依据视频文件的大小和网络最大传输单元 MTU(maximum transmission unit)来确定视频包的分包数量,将帧数量、分包数量等信息写入视频 IPB 帧末尾的用户数据域中(如图 16 所示),在接收端提取用户数据域内的信息并统计已正确接收到视频数据包的情况判断丢包的种类和数量.PLBU 算法利用 MPEG 4 或 H.264 视频流本身实现对分组丢失的测量,不影响视频的正常播放,测量引入的额外负载也较低,但算法目前仅可测量 MPEG 4 和 H.264 视频的丢包情况.

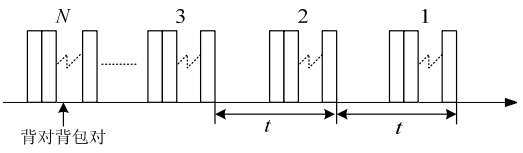


Fig.15 BADABING packet train[67]

图 15 BADABING 算法包列结构[67]

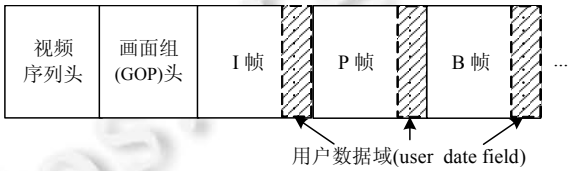


Fig.16 User data position in video[68]

图 16 用户数据域位置[68]

2.3 丢包模型

由于 IP 网络中丢包之间常存在一定的相关性,因此仅用丢包率还不能充分描述网络丢包特征.目前,评估网络分组丢失率的模型主要有贝努利(Bernoulli)模型、吉尔比特(Gilbert)模型、马尔可夫(Markov)模型[70]和双回归(double regression)模型[71]等.文献[72]通过长时间的实验发现:不同的采样间隔、不同的网络环境和不同的背景流量下,测量得到的分组丢失模型是不同的.由其所采集到的 38 组实验文件分析发现,有 7 个符合贝努利模型分布特性,10 个符合吉尔比特分布,21 个与 Markov 链模型相吻合.

© 中国科学院软件研究所 http://www.jos.org.cn

上述模型中,贝努利模型是基于独立同分布的,即,假定每个数据包在网络上传输时被丢弃的概率是不相关的.Gilbert 和 Markov 模型描述了连续丢包长度为 k(>0)的概率几何分布,它基于这样一个事实,即:如果数据包 n 丢失,那么 n+1 个分组丢失概率也比较高.Gilbert 模型是一种特殊的 Markov 模型(双状态 Markov 模型).Gilbert 模型中包括两种状态:“突发(burst)”状态描述的是分组突然大量丢失的情况;“间隙(gap)”状态描述了未发生分组丢失或随机的、互不相关的少量分组丢失的情况状态.Gilbert 模型状态转化如图 17 所示,图中,S_1 表示数据包丢失状态,S_2 表示分组正常接收状态,p_{11},p_{12},p_{22} 和 p_{21} 分别代表不同状态之间的转移概率,分组处于 S_1 的状态概率即为丢包率,$S_1=p_{21}/(p_{12}+p_{21})$,平均突发丢包长度(burst length distributions)为 $1/p_{21}$[72,73].近几年,Gilbert-Elliott 模型[74]和 4 阶 Markov 丢包模型[75,76]常被用来模拟网络实际丢包情况.4 阶 Markov 丢包模型状态转换如图 18 所示,S_2 表示分组正常接收状态,S_1 表示数据包丢失状态,p_{21},p_{12},p_{43},p_{34},p_{23} 和 p_{32} 代表不同状态之间的转移概率.

$$S_1 = 1\Big/\left(1+\frac{p_{12}}{p_{21}}+\frac{p_{12}\times p_{23}}{p_{21}\times p_{32}}+\frac{p_{12}\times p_{23}\times p_{34}}{p_{21}\times p_{32}\times p_{43}}\right),$$

$$S_2 = 1\Big/\left(1+\frac{p_{21}}{p_{12}}+\frac{p_{23}}{p_{32}}+\frac{p_{23}\times p_{34}}{p_{32}\times p_{43}}\right),$$

$$S_3 = 1\Big/\left(1+\frac{p_{34}}{p_{43}}+\frac{p_{32}}{p_{23}}+\frac{p_{21}\times p_{32}}{p_{12}\times p_{23}}\right),$$

$$S_4 = 1\Big/\left(1+\frac{p_{43}}{p_{34}}+\frac{p_{32}\times p_{43}}{p_{23}\times p_{34}}+\frac{p_{21}\times p_{32}\times p_{43}}{p_{12}\times p_{23}\times p_{34}}\right).$$

S_1+S_3 为总的丢包率, $\dfrac{S_1+S_3}{S_2(p_{21}+p_{23})+S_4(p_{43})}$ 表示平均突发丢包长度.

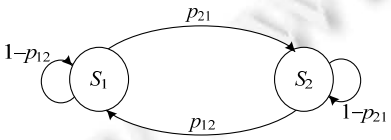


Fig.17 Gilbert model[72]

图 17 Gilbert 丢包模型[72]

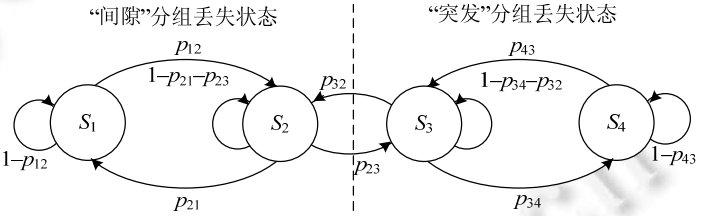


Fig.18 Four-State Markov chain[75]

图 18 4 阶马尔可夫链[75]

多数测量算法是在所选定的网络端点间进行端到端性能参数的测量,不能反映网络内部链路状态.目前,如何依据端系统间的测量结果推测出网络内部链路或节点的状态成为一个研究重点[77,78],其对于判断网络性能瓶颈、保证网络服务质量起着重要的作用.目前,这方面的研究成果主要集中于丢包测量领域[79],主要思想是:通过受控网络节点(通常是网络边缘的节点)收集端到端网络间有限的测量数据集,采用相似性和近似性的度量方法,利用统计推断的思想估计网络内部链路的性能状况.

2.4 应用流丢包与探测流丢包

现有丢包测量方法多为主动测量,即:通过测量包列的丢包信息来推测网络的丢包特性,进而估计应用流或背景流的丢包.但探测流丢包、网络丢包、应用流(背景流)丢包之间的关系较少有文献进行分析和说明.文献[67]最早阐述这个问题,并以是否与背景流丢包相一致来检验 BADABING 算法的性能.目前,大多数研究为了更好地分析数据包丢失情况,认为探测流(probe traffic)的丢包率就是背景流(cross traffic)的丢包率,且假设注入的测量数据包不引起背景流额外的丢包.从文献[67,68]的实验结果来看,应用流丢包与探测流丢包尽管存在较大的相关性,但两者之间仍有明显的区别(如图 19(a)所示).

假设探测流发包总数为 n_{pt},丢包数量为 n_{pl};背景流发包总数为 n_{ct},丢包数量为 n_{cl},则在该网络中的平均丢包率为$(n_{pl}+n_{cl})/(n_{pt}+n_{ct})$,要使探测流丢包、网络丢包、应用流(背景流)丢包情况相一致,那么探测流发包总数、丢

© 中国科学院软件研究所 http://www.jos.org.cn

包数量就必须与背景流的发包总数、丢包数量相等.在实际网络中,上述条件显然难以满足.文献[67,68]的实验也佐证了我们的论述.文献[68]对探测流丢包、网络丢包、应用流(背景流)丢包情况进行了分析,图 19(b)所示为多组实验中探测流丢包、网络丢包、应用流丢包最为接近的一组.本文认为,丢包测量算法存在的意义就是要通过对网络路径的探测来反映应用流或背景流的丢包状况.目前的测量算法中,仅 PLBU 算法实现了探测流与应用流丢包的一致性,但 PLBU 算法仅针对特定的视频,应用范围极为有限.如何使得测量到的丢包情况与应用流的丢包情况相一致或相近,是丢包测量领域重要的研究方向.

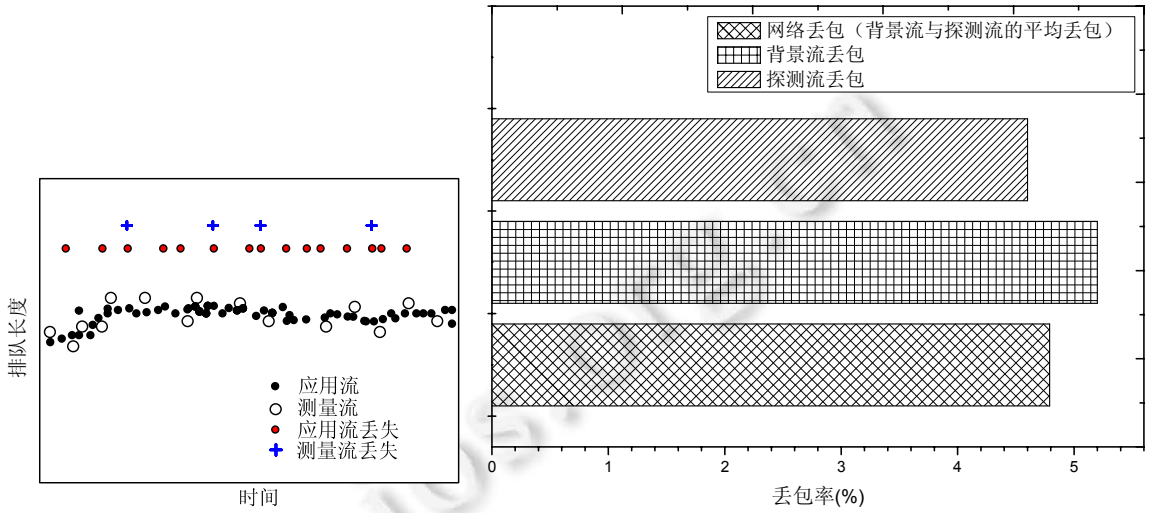


Fig.19 Difference between measurement flow and application stream packet loss[67]

图 19 测量流与应用流丢包之间的差异[67]

2.5 算法对比分析

文献[67]的实验结果表明:同等测量负载的情况下,BADABING 算法精度优于 ZING 算法.而 PLBU 算法借助视频流自身完成对丢包的测量,直接测量应用流丢包情况,这与传统方法完全不同.因此,以 BADABING 算法和 PLBU 算法作为分析对象.

实验拓扑如图 20 所示,其中,视频服务器选用 Helix Server,网络模拟软件 Nistnet 用来产生网络路径丢包,网络嗅探工具 Wireshark 用来抓取网络全部流量,进而统计应用流实际丢包情况.

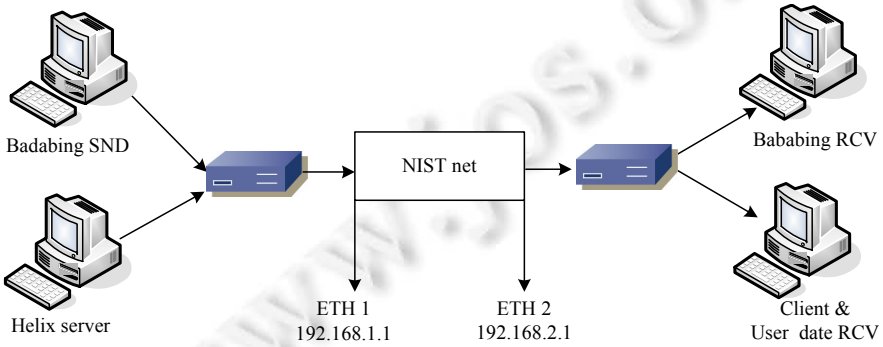


Fig.20 Tested configuration

图 20 实验拓扑

实验结果显示(如图 21 和表 2 所示):BADABING 算法测量结果与背景流的丢包情况存在明显的差异,表明

© 中国科学院软件研究所 http://www.jos.org.cn

探测流的丢包并不完全与背景流的丢包相一致;而PLBU的测量结果却与背景流的实际丢包相一致,PLBU算法将视频帧转化为测试序列,通过在视频文件 User_Data 域中嵌入特定的测量信息完成对丢包的测量,这种测量方法不生成任何新的测试流,测量结果就是应用流的具体丢包情况.尽管目前 PLBU 算法思想仅可应用于MPEG 4 和 H.264 视频,但我们认为,这种研究思路为提出新的测量算法提供了借鉴.更加详细、完整的丢包测量算法对比分析结果可参考文献[67,68].

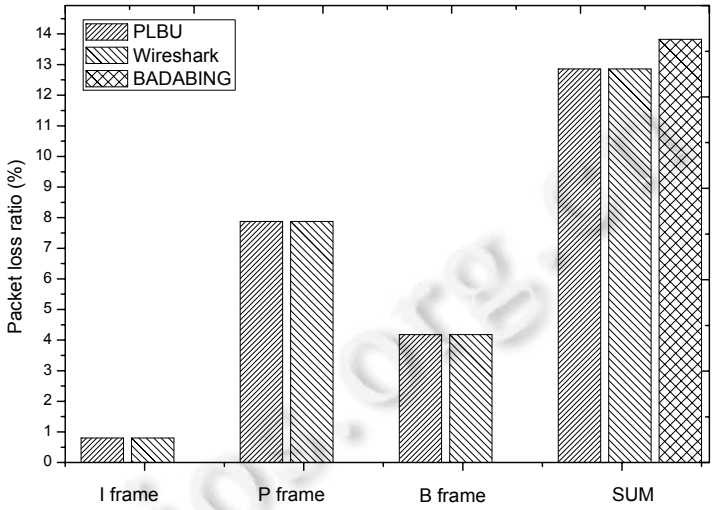


Fig.21 Experimental results

图 21 实验结果

Table 2 Experimental results

表 2 测量结果

测量方法	I 帧丢包率(%)	P 帧丢包率(%)	B 帧丢包率(%)	总丢包率(%)
Wireshark	0.803 8	7.877 8	4.180 0	12.861 7
PLBU	0.803 8	7.877 8	4.180 0	12.861 7
BADABING	–	–	–	13.823

3 时延相关测量

时延通常分为往返时延和单向时延,并由其衍生出时延变化(抖动)等网络性能参数.测量往返时延时,由于测量的开始与结束时间都由测量发起端时钟记录,简单易行,也无需时钟同步算法的引入.但在实际使用中,单向时延指标往往更能准确反映网络向应用实际提供服务的质量等级,如视频点播服务,视频数据流通常是在服务器端到客户端的方向上传输.根据 RFC 2679 的描述,网络单程延迟的测量需要建立在路径两端主机时钟同步的基础之上[80].此外,带宽测量时常依据时延的变化情况来判断算法是否收敛,如 pathload 算法、SLDRT 算法等;时延对丢包也有着直接影响,如文献[81]的研究表明,时延值变化与丢包之间存在着直接关系.

3.1 基本概念

定义 10(单向时延(one way delay,简称 OWD)). 分组从发送端发出到接收端收到时所经历的时间差[80].

定义 11(往返时延(round trip time,简称 RTT)). 从发送方发出数据包开始,到发送方收到来自接收方的确认(接收端收到数据后便立即发送确认消息)所经历的时间差[82].

定义 12(抖动(jitter)). 抖动又称为时延变化,是指分组时延的变化程度,主要用于表征以相同时间间隔发送的数据包却以不规则的时间间隔到达接收端的现象[83].

定义 13(时钟偏差(clock offset)). 这是指在某一时刻时钟时间与参考时间之间的误差[84].

© 中国科学院软件研究所 http://www.jos.org.cn

设 $C_S(t)$,$C_R(t)$分别为在 t 时刻(t 为标准时间 UTC(coordinated universal time))发送端(sender)时钟和接收端(receiver)时钟,则 C_S 的时钟偏差为 $C_S(t)-t$,C_R 的时钟偏差为 $C_R(t)-t$.假设时钟 C_S 为参考时间,那么 $C_R(t)$相对于 $C_S(t)$的时间偏差为 $C_R(t)-C_S(t)$.

定义 14(时钟频率(clock frequency)). 时钟频率反映了时钟时间变化的快慢程度,时钟 C 的频率 $C'(t)$可定义为 $C'(t)=\mathrm{d}C(t)/\mathrm{d}t$.

定义 15(时钟频差(clock skew)). 一个时钟的频率与参考时钟频率之间的差值称为时钟频差.假设时钟 C_S 为参考时钟频率,时钟 C_R 相对于 C_S 的时钟频差为 $C'_R(t)-C'_S(t)$[84].

消除端系统间的时钟偏差是实现网络同步、进而实现单向时延测量的前提条件.而由于时钟频差的客观存在,又使得时钟偏差在不断地发生变化.因此,与时延测量算法相关的研究常划分为时钟偏差测量和时钟频差测量两类.

3.2 时钟频差测量

文献[84−86]将时钟频差消除问题归结为寻找一个线性函数的过程,即:对于测量集合 $\Omega=\{v_i=(t_i,d_i), i=1,\ldots,N\}$(表示 t_i 时间测得的单向延迟数据 d_i),求一个线性函数,该函数满足以下条件:

(1) 所有(t_i,d_i)在所求的分段函数上方(单向时延没有负值);

(2) 线段最接近测量集合Ω.

考虑到计算机中产生时间中断的石英晶体振荡器通常较为稳定,计算机的时钟频率通常为定值,则可假设时钟频差的曲线是 $L:=\{(x,y)|y=\alpha x+\beta\}$,条件(1)就可以表述为 $d_i\geqslant\alpha x+\beta$.

对于所有满足以上条件的直线,选取最接近Ω的那条(即条件(2)).

这个问题归结于一个最优化问题,它应从 3 个指标考虑,从而有如下 3 个目标函数.

1) 所有采样点与直线的垂直距离之和最小化.

$$obj_1:=\sum_{i=1}^{N}(d_i-\alpha t_i-\beta)=\sum_{i=1}^{N}d_i-\sum_{i=1}^{N}\alpha t_i-N\beta \tag{9}$$

2) 采样曲线与直线之间的区域面积最小化.

$$obj_2:=\sum_{i=1}^{N}(d_i-at_i-b+d_{i+1}-at_{i+1}-b)\frac{(t_{i+1}-t_i)}{2}=\sum_{i=1}^{N}\frac{(d_i+d_{i+1})(t_{i+1}-t_i)}{2}-\frac{t_N^2-t_1^2}{2}a-(t_N-t_1)b \tag{10}$$

对于采样时间间隔相等的特例,即 $t_{i+1}-t_i=c$,有:

$$\frac{N-1}{N}obj_1-\frac{1}{c}obj_2=\frac{d_1+d_N}{2}-\frac{d_1+\ldots+d_N}{N} \tag{11}$$

对给定的一组数据,由于 c,N和 d_i 都是常数,因此 obj_1和 obj_2这两个目标函数是等价的(可以互相转换).

3) 使落在直线上的点最多.

$$obj_3=\sum_{i=1}^{N}1_{\{d_i=\alpha t_i+\beta\}} \tag{12}$$

时钟频差的消除问题可转化为找到一条满足限制条件的直线,使以上 3 个目标函数最大化的问题.

为了求解这一问题,文献[85]采取分段最小值的方法,首先确定各个分段中最小值点,再把这些点连成一系列直线段,利用这些线段来估计时钟频差.但这种方法在网络拥塞状况严重、排队时延持续较高的情况下准确性和稳定性较差.文献[84]提出了 LPA 算法(linear programming algorithm),利用回归分析方法试图找出一条倾斜直线,使单向时延的测量值到该直线的距离最小化,从而达到测量时钟频差的目的,但算法计算复杂度较大.

文献[86]提出了 CHA 算法(convex hull algorithm),把约束条件转化为求时延数据全集Ω的凸包问题:

$$co(\Omega):=\left\{x\middle|x=\sum_i\lambda_i v_i,\lambda_i\geqslant 0,\sum_i\lambda_i=1,v_i\in\Omega\right\},$$

将求解目标函数的解归结为如何求凸包集下边界的问题.CHA 算法可在线性时间内解决问题,这是一种较为准确、快速的方法.

© 中国科学院软件研究所 http://www.jos.org.cn

文献[87]提出一种基于模糊聚类分析的算法来检测时钟频差,采用最小滤波方法去除数据噪声,使用线性规划或凸包方法估计全局最优的时钟频差值.文献[88,89]提出利用CHA算法来消除时钟频差,同时用Intel CPU内部时间戳计数器TSC(time stamp counter)取代系统时钟计时来提高软件时钟分辨率,以消除测量时钟的计时误差.虽然这一方法具有精度高、开销小的特点,但TSC寄存器只是Intel奔腾系列CPU独有的64位计数器,该方法目前无法应用于使用其他CPU型号的系统中.文献[90]利用图像处理中的Hough变换投票过程(Hough transform voting process)将时延值分布限定在一个平行四边形区域,平行四边形的底边即为计算时钟频差所需确定的直线.与基于线性规划的LPA算法相比,该方法具有较好的稳定性.此外,文献[91]提出,通过间歇GPS信号周期性地消除系统间的时钟频差.

上述方法仅考虑了时钟频差的估算问题,而没有与时钟偏差测量相结合.尽管测量中可先获得时钟频差,再通过引入外部时钟源的办法来获得时钟偏差值,最后再进行单向时延的测量,但这种方法使用上极不方便.因此,仅用上述时钟频差测量方法,还不能完全解决测量中的时钟不同步问题.

3.3 时钟偏差测量

借助GPS[92]、NTP(network time protocol)[93]和精密时钟同步PTP协议(IEEE 1588 precision clock synchronization protocol)[94]来测量和消除时钟偏差是目前的主流方法.如文献[95]以GPS时钟为基准时钟源,采用一个称为时钟桥的时钟中继装置对操作终端进行授时,但借助GPS等外部时钟源,虽可高精度地消除时钟偏差,但是这种方式所需硬件价格昂贵且与接收环境有关,难以大规模地推广运用.NTP的同步机制发送一个类似PING的探测包,该包携带发送端发送数据的开始时间戳,接收端在收到探测包后返回一个应答包,应答包携带探测包的发送时间戳、探测包的接收时间戳和应答包的发送时间戳.根据这几个时间戳,NTP计算出两台机器的时钟偏差,从而完成同步.基于NTP同步方法的误差极限小于*RTT*/2,NTP时钟同步的准确性则依赖于同步主机间的路径时延,在假设往返路径时延相等的前提下才能获得理想的测量结果.IEEE 1588 PTP协议借助硬件设备将网络设备的时钟与主控机的主时钟实现同步,同步过程中,主时钟周期性地发布时间同步协议及时间信息,系统据此计算出主从线路时延及主从时间差,并利用该时间差调整本地时间,使主从设备时间保持一致来消除时钟偏差.但该方法不支持非对称路径下的时钟同步,也需要特殊硬件设备的支持.文献[96]利用卡尔曼(Kalman)滤波对IEEE 1588时钟算法进行了改进,对同步算法中的噪声和干扰因素作了进一步剔出.文献[97]利用已有时钟偏差值和对过去时钟偏差值的指数平均来估计当前时钟偏差的变化情况,并与NTP和IEEE 1588算法获取的时钟偏差情况进行了对比分析.该方法虽可获得较为理想的精度,但同样不适用于非对称路径(asymmetric paths)下消除时钟偏差的处理.

此外,随着各类新型网络结构的不断出现及网络规模的迅速扩大,网络路径中非对称路径已占得相当比例,依托NTP和PTP协议均难以在非对称的网络路径中获得精确的时钟偏差,进而测量得到准确的单向时延.文献[98]提出一种非对称路径情况下的时钟偏差测量算法,其理论误差极限小于*RTT*/4.但该算法中约简变量的前提假设是往返路径中的抖动相同,显然,这个假设条件在网络中不可能总是得到满足.由于在非对称路径中,传统的依据测量包的时戳信息来计算时钟偏差的办法已不可行,仅通过简单分析测量包的时戳信息已难以准确测量到时钟偏差,因此,一些研究人员开始尝试在不消除时钟偏差的前提下进行单向时延及其他相关参数的测量,但成果亦有很大的局限性.如:文献[99]研究了单向时延抖动的测量方法,但该方法必须利用Windows NDIS(network driver interface specification)来消除测量时钟的计时误差;文献[100]提出了改进的最大熵目标函数来测量单向时延,但该方法对网络链路和循环路径的数量有严格要求,并且事先需获得多个节点对之间的时延值,且不能对单条路径进行针对性的测量;文献[101]利用排队时延通常具有伽玛(Gamma)分布的特性,采用预先设定时延分布函数,并结合分位数图(quantile-quantile图)的办法来获取排队时延为0情况下的最小测量时延值(即时钟偏差、传输时延、传播时延之和).该方法提供了一种新的最小时延测量方法,对于如何在网络背景流分布特性不明确的情况下测量最小时延提供了借鉴;在传输时延与传播时延已知的前提下,文献[101]给出的方法可准确获知节点间的时钟偏差,但在实际网络中,网络传输时延与传播时延通常难以获得,显然,该方法适用范围极为有限.尽管文献[101]中提到该方法适用于非对称路径,但在理论和实验上均未给出相应的说明或证明.

© 中国科学院软件研究所 http://www.jos.org.cn

3.4 时钟偏差与时钟频差

时钟偏差与时钟频差是实现网络同步的主导因素,两者之间存在相互关联的关系.时钟频差主要由计算机内石英晶体振荡器产生的中断所决定,通常情况下保持稳定,但低精度振荡器也可使时钟频差发生变化[102].由于不同计算机内产生中断的石英晶体不可能完全相同,受温度、湿度和气压差异等诸多因素的影响,不同计算机的时钟会以不同的时钟频率走出时钟滴答[103],从而导致随着时间的增加,端系统之间的时间偏差呈增长/下降趋势.可以说,时钟偏差受时钟频差的影响,并不总能保持为常量.

在已知时钟频差的条件下,即:发送端时钟频率 $C'_S(t)$,接收端时钟频率 $C'_R(t)$ 已经确定.以发送端时钟为参考时钟,设记录时钟偏差的起始时刻为 t_0,则此时发送端时钟为 $C_S(t_0)$,接收端时钟为 $C_R(t_0)$.由时钟频率定义可知:经过Δt 时间间隔,发送端、接收端时钟所走过的时间长度为 $C'_S(t)\times\Delta t, C'_R(t)\times\Delta t$.即:经过$\Delta t$ 时间间隔,发送端时钟变为 $C_S(t_0)+C'_S(t)\times\Delta t$,接收端时钟变为 $C_R(t_0)+C'_R(t)\times\Delta t$,则此时时钟偏差变为

$$(C_R(t_0)+C'_R(t)\times\Delta t)-(C_S(t_0)+C'_S(t)\times\Delta t)=(C_R(t_0)-C_S(t_0))+(C'_R(t)-C'_S(t))\times\Delta t.$$

由上式可知,$t_0+\Delta t$ 的时钟偏差由 t_0 时刻的时钟偏差$(C_R(t_0)-C_S(t_0))$和由时钟频差引起的时钟偏差变化量$(C'_R(t)-C'_S(t))$两部分构成.通常可认为 $C'_S(t)$ 和 $C'_R(t)$ 为常数.随着时间的增长,时钟偏差会越来越大.如果在测量中不考虑时钟频差的影响,会对时钟偏差的测量结果产生积聚性误差.设 $C'_R(t)$ 大于 $C'_S(t)$,上述分析结论的图形化表示如图 22 所示.

图 22 中,斜线表示时钟偏差随时间的变化情况.在实际网络环境下,我们对上述分析进行了验证.图 23 为两台未经时间同步的计算机之间测量出的时延变化情况.实验中:一台计算机作为发送端,每秒发送一个打上发送端发送时间戳的探测包;另一台计算机作为接收端,接收端在接收探测包时记录下接收时间戳,接收端计算探测包发送时间戳与接收端接收时间戳的差值(接收时间 T_R 减去发送时间 T_S,记为 T_R-T_S,显然,$\Delta offset=T_R-T_S$).受程序发包和机器性能等因素的影响,虽然测量到的时延值呈现出一定程度的波动,但时钟偏差变化趋势与图 21 所示的理论分析相一致.

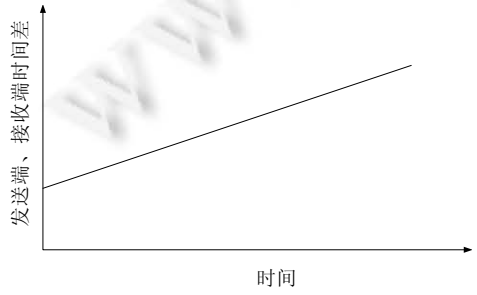


Fig.22 Change of clock skew (theoretical value)

图 22 时钟偏差变化(理论值)

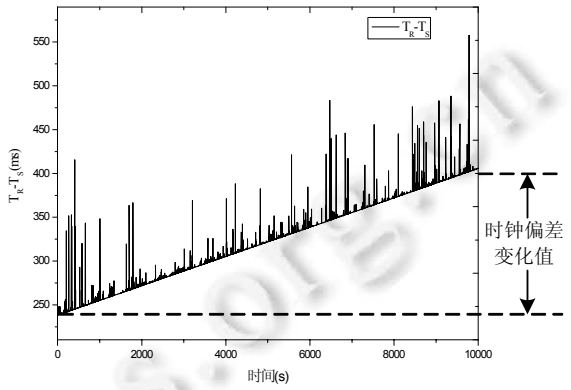


Fig.23 Change of clock skew (actual value)

图 23 时钟偏差变化(实际测量值)

当前,已有的研究成果较少分析时钟频差与时钟偏差共同作用情况下如何实现时钟同步的问题,也还未提出误差精度在可应用范围内的非对称路径条件下时钟偏差测量算法.

3.5 算法对比分析

单向时延的测量是时延测量领域的难点,针对单向时延的测量,需要在已知或消除系统间的时钟偏差的基础上来进行.

本文分别选用对称路径下的 NTP 方法和适用于非对称路径的 KIM 算法[98]来说明不同时钟偏差测量算法的性能.在不引入外部时钟源的情况下,本文用 NTP 或 KIM 算法得出时钟偏差,完成系统同步后,再测量单向时

© 中国科学院软件研究所 http://www.jos.org.cn

延.实验拓扑如图 24 所示.在对称路径情况下,往返路径的带宽均设为 0.1Mpbs;在非对称路径情况下,往返路径带宽分别设为 0.1Mpbs 和 1Mbps.在实验室环境下,因传输距离较短,传播时延可忽略.本文以数据包的传输时延作为单向时延准确值来分析测量误差.传输时延计算方法如下:设探测包的总大小为 1 042 字节(测量包大小为 1 000 字节,包头大小为 42 字节),在 0.1Mpbs 带宽环境下,传输时延值为(1042×8)/100000=83.36ms;在 1Mpbs 带宽环境下,传输时延值为(1042×8)/1000000=8.336ms.

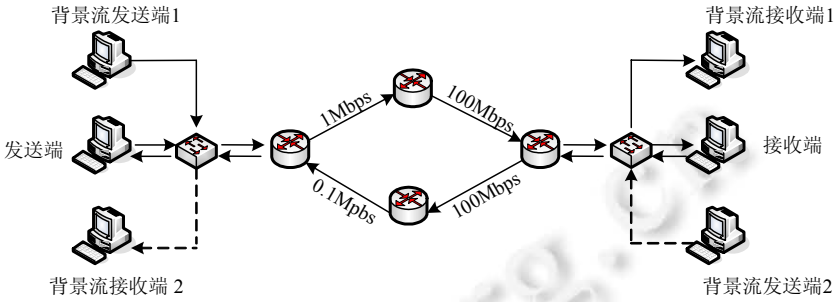
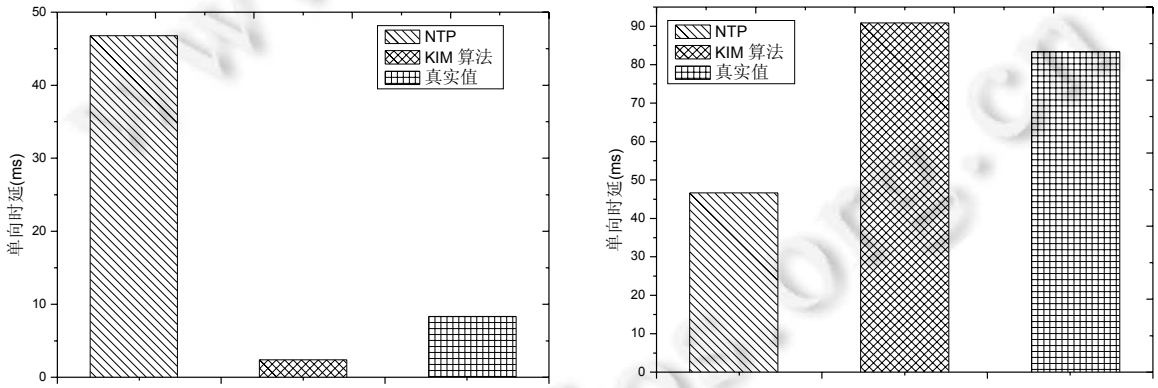


Fig.24 Tested configuration

图 24 实验拓扑

在对称路径情况下,依 NTP 算法实现系统时钟同步,往返路径中单向时延测量值分别为 84.028ms,83.915ms;KIM 算法测量值分别为 83.983ms,84.059ms.从实验结果来看:虽然受主机性能波动的影响,测量结果有轻微的变化,但总体与理论值非常接近.

在非对称情况下,依 NTP 算法实现系统时钟同步,往返路径中单向时延测量值分别为 46.770ms,46.636ms;KIM 算法测量所得不同方向单向时延值分别为 90.878ms,2.395ms.实验结果表明:在非对称路径下,NTP 算法的测量结果已远离真实值.实验结果如图 25 所示.



(a) 从 SND 到 RCV 的单向时延 (b) 非对称路径下 RCV-SND 单向时延

Fig.25 Experimental result on asymmetric paths

图 25 在非对称路径上时延测量结果

理论分析可证明:NTP 算法在非对称路径下测量误差上界为 *RTT*/2,而 KIM 算法测量误差范围上界为 *RTT*/4[98].这样的误差范围显然并不足以有效地消除时钟偏差,从而实现高精度网络时钟同步.因此,方便、准确的非对称路径下的时钟偏差算法尚待研究.

© 中国科学院软件研究所 http://www.jos.org.cn

4 总结和展望

4.1 总 结

对带宽、丢包、时延的测量是量化分析网络性能的基础性工作,测量算法的准确性、稳健性对协议设计、QoS 部署、多媒体传输等诸多方面的研究有着直接影响.从 1988 年第一个链路带宽测量算法 pathchar 提出至今已有 20 多年的时间,拥有了数量众多的各类算法,许多代表性的算法也已得到了广泛应用,但仍有诸多关键问题尚待解决.

4.1.1 带宽测量算法存在的问题和挑战

(1) 背景流突发性对测量的影响

绝大多数带宽测量算法基于以下假设前提:① 背景流满足流体模型(设背景流发送速率为 x,在任意时间间隔 t 内,到达某一链路的背景流量均为 $x\times t$);② 背景流在单个测量周期中保持恒定.显然,这个假设条件与网络实际流量的突发性特征相矛盾.本文在第 1.5 节分析了突发背景流对带宽测量可能带来的影响.要研究背景流对测量算法的影响,首先需要用数学模型将背景流特征进行表征,但当前,流量模型(主要有 Poisson 分布模型、Deterministic 分布模型、Pareto 分布模型、Long range dependent 模型)与实际网络流量的分布特性均存在较大差距.因此,目前带宽测量算法在突发背景流下的正确性和可行性还主要是以实验验证为主,因而导致带宽测量算法的健壮性始终无法从理论上完全获得证明和保障.文献[61,62]的研究指出:在突发背景流下,完全准确的测量可能难以做到,但长的探测包列会减少测量误差.这为高精度带宽测量算法的研究提供了一个方向.

(2) 路由器队列管理对测量的影响

带宽测量算法根据测量包对(或包列)经过窄链路(或紧链路)之后,包对间隔或包列速率的变化情况来判断带宽值.显然,窄(紧)链路处的路由器队列管理方式会对测量包的间隔和速率变化产生影响.尽管多数算法假设待测路径路由器采用先进先出(FIFO)的路由器排队,但在实际网络中,情况并非如此.目前,为了避免网络拥塞,减小排队时延,保证较高的吞吐量,越来越多的路由器采用了主动队列管理(active queue management,简称 AQM 算法)方式.与 FIFO 相比,AQM 通过对拥塞的预判断和主动丢包,实现对网络拥塞的控制.AQM 算法对带宽测量带来的主要问题是:

① 探测流为测量带宽而故意造成的网络拥塞可能会因 AQM 算法的存在而无法实现,使测量算法难以达到收敛条件;

② 由于 AQM 算法为避免网络出现拥塞会主动丢弃数据包,这将导致以数据包丢失为网络拥塞判断条件的测量算法产生错误判断,即:当丢包出现时,算法提前认为探测包对或包列速率已高于待测带宽的阈值,最终得出不正确的测量结论;

③ AQM 算法可能会使经过窄(紧)链路的测量包对或包列结构发生变化,而这种变化并非由背景流量造成,这使得探测包对间隔或包列速率变化不能正确反映出探测包速率、背景流量和带宽之间的关联关系,进而造成错误判断,产生测量误差.

目前,关于路由器采用主动队列管理时带宽应如何测量的问题,还较少有文献进行研究.

4.1.2 丢包测量算法存在的问题和挑战

(1) 探测流与背景流差异问题

丢包测量中,通常把探测流的丢包当作背景流(应用流)的丢包,进而用其来表征当前网络的丢包情况.如果探测流丢包与网络总体丢包情况相一致,探测流发包总数、丢包数量就必须与背景流的发包总数、丢包数量相等.在实际网络环境下,上述条件显然难以满足(见第 2.4 节分析).网络中,通常使用设置优先级队列的方法来优先保证视频等应用的服务质量.这种情况下,探测包与应用数据包的优先等级将会不一致.此时,若再使用 BABABING,ZING 算法的测量结果去判断应用流的丢包情况,则测量误差还会加大.如果能将应用流转化为测量包列,利用应用流自身直接来测量丢包,则从根本上解决了丢包测量中探测流与背景流的差异问题,对于提高测量精度、减轻测量负载具有重大意义.

© 中国科学院软件研究所 http://www.jos.org.cn

(2) 丢包模型在线分析问题

当前,丢包测量算法主要关注对平均丢包率的测量,但研究表明:丢包率并不能完全表征网络丢包特性,丢包的时空分布情况对网络应用流的影响同样不可忽视.丢包模型目前需通过离线分析探测流或背景流的 trace 文件才能获得.在丢包测量算法中,BADABING 算法可以测量出丢包发生的持续时间,PLBU 算法可以测量出每个视频帧的丢包数量,但 BADABING,PLBU 等算法均无法在线获知每个丢包发生的具体位置,因此并不能依据测量结果实时地判断出丢包发生的时、空间关系,进而分析出与何种丢包模型相匹配.另外,丢包模型目前还缺乏统一定义,不同的研究者提出了不同的模型(Bernoulli,Gilbert,Gilbert-Elliott,Markov,Double Regression 模型,等等),从而增加了解决丢包模型在线分析和运用问题的难度.

4.1.3 时延测量算法存在的问题和挑战

(1) 时钟频差与时钟偏差的相互作用问题

时钟频差对时钟偏差的作用影响,在本文的第 3.4 节已进行了形式化分析,此处不再赘述.计算机时钟频差主要由其石英晶体振荡器产生的中断所决定,由于计算机石英晶体振荡器不可能完全相同,因此计算机的时钟频差是客观存在,且两两各异.若要用软件的方式实现计算机间高精度的时钟同步,就必须考虑时钟频差对时钟偏差的影响,这不可避免地会增加算法设计的难度.

(2) 非对称路径下时延测量问题

往返时延(RTT)测量由于测量开始与结束时间都由测量发起端时钟记录,因此无需测量端系统间的时钟偏差.而单向时延的测量就必须建立在时钟偏差已知的基础上.现有时钟偏差测量算法主要针对对称的网络路径,但是随着互联网规模的不断扩大,网络结构日益复杂,非对称路径(asymmetric paths)已占网络路径的相当部分(如互联网技术全球性测试平台 PlanetLab 网络中非对称路径已占其全部路径的 10%~15%).如果测量路径是非对称性的,测量包的时间戳信息与时钟偏差关系就会变得复杂,建立起测量包时间戳、时钟偏差和单向时延之间的函数关系也会更加困难.

4.1.4 网络性能测量面临的共性问题和挑战

除上述问题以外,网络性能测量还面临以下共同的问题和挑战.

(1) 计时精度对测量的影响

带宽测量、时延测量算法要求精确控制测量数据包的发送时间和发送间隔,并要求准确获取测量包的发送和接收时间戳信息.因此,系统的计时精度对测量结果有重大影响.在实际网络环境下,线程切换、网络中断技术、系统调用延迟(包括获取系统当前时间、读/写套接字时间)均可能影响计时精度,并带来计时误差[21].由于 Window,Linux 等通用操作系统目前只能提供毫秒级的计时器,这对于需要精确计时的测量系统显然是不够的.以带宽测量为例,在 Window,Linux 此类通用操作系统上,pathChirp,pathload 算法通常能在 10/100 兆网络环境中获得较为理想的测量结果,而在高速网络环境下(1G 带宽以上)却无法得到正确的测量结果.这是因为,即使取最大测量包,要实现对百兆以上的网络进行测量,探测包最小时间间隔应精确至微秒级(μs),但普通操作系统无法保证在此条件下的计时精度.文献[104]利用 FPGA(field programmable gate array)等手段,通过在网络中加入 middlebox 的方法实现了对高速网络的测量.文献[89]利用 Intel CPU 内部时间戳计数器 TSC(time stamp counter)提高了时钟分辨率.这些方法为如何在通用平台上获取高精度的计时信息提供了解决方案,但必须借助部分硬件设备和信息,适用范围仍受到限制.

(2) 动态路由问题

目前,互联网采用的是动态路由协议,即:互联网中任意两点间的路由是动态变化的,节点间的数据包在传输过程中可能经过不同的网络路径.路由的不确定性可能导致探测包流量途经路径的属性发生改变,即,背景流量、带宽大小、传输媒介在测量过程中发生了变化,进而导致测量失败.文献[105]研究表明:网络路由的变化以小时为单位,通常可以保持稳定状态.我们对 Planetlab 的路由变化情况进行了分析,实验结果表明:在 10~30 分钟的时间内,路由基本上可保持稳定.虽然当前的测量算法在秒级的时间均可得到测量结果,所以假设在测量过程中路由不发生变化是合理的,但对网络进行持续及大规模测量时,动态路由问题就必须要作为一个因素而加以

© 中国科学院软件研究所 http://www.jos.org.cn

考虑.

(3) 测量负载问题

主动测量算法因其灵活性而成为当前网络性能测量算法中的主流,但主动测量需要向网络注入测量包,会对背景流产生负面影响.此外,为了降低网络噪声对测量的影响,各种算法均选用较大的测量包(如:pathChirp 测量包大小为 1 000B,SLDRT 测量包大小为 1 200B,BADABING 测量包大小为 600B,等等),这更加重了因测量而引入的额外负载.我们通过实验发现:带宽、丢包、时延测量引入的测量负载通常以 MB 为单位,这对百兆或千兆网络影响不太大,但在低带宽网络环境下,运用这些算法就会给网络带来极大的负担(如:BADABING 算法测量丢包时,占用带宽为 0.864Mb/s).如何降低测量负载,就成为一个重点问题.文献[68]利用视频流中的用户数据域(user-data field)实现对丢包的测量,文献[106]利用 TCP/IP 的边信道(side channel)实现了对 RTT 的测量.这些研究通过挖掘和利用背景流量的特性,依托背景流自身即可实现对网络性能的测量,极大地减轻了测量负载,也为其他测量方法的研究提供了借鉴.

(4) 最小时延测量问题

如果时延测量值中不包含数据包在路径中因突发背景流量而造成的排队等候时间,则该时延被称为最小时延.最小时延在带宽和时延测量领域均有广泛运用,且对测量精度影响很大,如,pathChar,pchar,TRIO[107],GeoPing[108],CapProbe 等算法的测量准确性均由最小时延决定.但是如何测量最小时延,目前还没有得到业内一致认可的研究成果,不同的测量算法发送数量不等的探测包来获取最小时延(如:pathChar 测量包数量为 32 个,CapProbe 测量包数量为 100 个,GeoPing 测量包数量为 10~15 个,TRIO 测量包数量为 3 个).由于最小时延测量的重要性,一些文献对如何测量最小时延进行了初步分析,如文献[29]分析了不同背景流模型下测量最小时延所需的测量包数量,文献[109]提出用最小时延区分 MDDIF(minimum delay difference)的方法来从测量数据中提取最小时延.但要从理论上保证最小时延的可测性,就需要从背景流量的分布特征来入手.由于针对实际网络背景流的数学模型目前还没有权威的研究成果,导致如何保证能测量得到最小时延、如何证明测量得到的最小时延是正确的等许多方面的研究还存在空白.目前,利用核密度估计函数、分位数图、卡尔曼滤波、聚类分析方法[110]等统计方法来从测量数据中有效地过滤出最小时延,可能是一个研究方向.

4.2 展 望

综上分析,结合计算机网络的最新发展,网络性能测量在以下几个方面还有待进一步的研究.

4.2.1 测量算法的进一步完善

尽管网络性能测量研究成果众多,但并不意味着算法已完全趋于成熟.综合第 1.5 节、第 2.4 节、第 3.4 节和第 4.1 节的分析,我们认为:在测量算法存在的诸多不足中,突发背景流条件下的带宽测量算法、利用应用流自身数据的丢包测量算法、适用于对称和非对称路径下的时钟偏差测量算法是下一步研究的主要方向.此外,当前的测量方法通常以测量精度作为最重要的评价指标,专注于对某一参数的测量,要获得多个网络性能指标,就必须部署多个不同的测量工具,这样的测量方式需向网络注入大量的测量包,测量包需反复穿越网络路径,测量时间较长,不可避免地会增加网络负载,进而可能改变网络的运行状况.并且,网络中的具体应用同时会受到多个性能参数的影响,单个性能指标的获取并不能完全表征网络现状.若能在一个测量周期或较短的时间内同时获取多个网络性能参数值,将会极大地推动网络性能测量领域的研究.虽然不同算法的基本思想存在巨大的差异,但也应看到:自拥塞算法思想、背靠背探测包对等测量技术,在丢包测量、带宽测量、时延测量等多个领域均有应用.是否能适当降低测量精度要求,以各种算法采用的共性技术为基础设计出多网络性能参数同时测量算法,应是研究的关注点之一.

4.2.2 网络测量任务部署算法

互联网已是一个复杂的巨系统,且规模还在不断扩大,网络有限测量资源与多样化测量需求之间的矛盾也日趋凸显.要想获取全网络或部分网络的性能信息,而去测量网络中每个链路或路径显然是不可行.除提高算法的测量效率之外,把测量节点部署在最合理的位置,减少因测量而引入的附加流量,提高测量覆盖率,也是一个重要的研究方向.文献[111,112]对该问题进行了较为系统的阐述和分析,总结了网络测量部署模型目前存在的

© 中国科学院软件研究所 http://www.jos.org.cn

不足,指出,挖掘网络信息是提高测量部署模型和优化算法效率的有效途径.

4.2.3 测量信息的综合和建模

不同网络性能测量算法所得信息种类繁多、量纲各异,如时延单位通常是 ms,带宽测量单位通常是 Mbps,而丢包测量结果通常用百分数来表示,测量信息之间相互割裂,且彼此之间存在明显的异构性.目前的测量算法研究还没有将多个网络性能指标综合起来,给出一个可量化、易于理解的统一的表示形式.由于网络上流媒体应用的爆炸式发展,网络性能对这些新兴应用的影响已无法用单个性能参数值来表征,可以说,传统的评价指标已无法表征当前网络性能和应用服务的状态.一些研究者已认识到这个问题,开始将用户体验质量 QoE(quality of experience)引入到网络性能测量领域,如文献[113,114],虽然其研究的主要目的是分析网络性能变化对流媒体体验质量的影响,但其实验方法与分析手段为如何对网络运行状态做出恰当的、规范和统一的评价提供了借鉴.

4.2.4 网络测量信息挖掘

针对大规模网络的性能测量研究是近年来网络测量领域的关注热点之一[115,116].在当前的技术手段下,要实现对大规模网络的性能评估,需要长时间持续不断的测量,才能获取完整而准确的数据.因此,网络性能测量还需面对大规模测量可能带来的海量数据和数据时效性的问题.另外,通过对网络性能测量信息的跟踪、统计、评价、分析、优化,使管理者有可能迅速、直观地判断网络性能的变化趋势,并及时定位网络性能故障位置,甚至可能在网络性能出现劣变之前就予以解决,保证网络可访问和非拥挤,使传统网络测量为事后处理模式向预先评估预警模式转变.大数据分析、云计算等手段的出现,为实现网络性能的在线分析和预测提供了可借鉴的方法和手段,有着广泛的科学意义和应用价值.当前,清华大学已率先将大数据技术运用于网络性能分析和预测之中[117].

4.2.5 无线网络、下一代互联网性能测量算法

无线网络技术,特别是近年来 WiFi,3G/4G 技术的飞速发展,使得通过无线方式接入互联网已成为人们上网的一种主要方式,针对无线网络性能测量的研究也逐渐成为当前的研究热点[118].我们通过实验发现:现有测量算法大多只适用于传统有线网络环境,当测量算法被用于无线网络环境(或有线与无线并存的混合网络环境)时测量误差较大,有时甚至无规律可循.无线网络特有的信道特性、网络协议、移动随机性,进一步加大了测量难度.此外,近年来软件定义网络 SDN(software-defined networking)和命名数据网络 NDN(named data network)等新兴网络的出现,也为网络性能测量的研究提出了新的课题.如 NDN 网络已不再使用 TCP/IP 体系结构,则现有测量算法的适用性、准确性均面临极大的挑战,但也为研究者提供了巨大的探索空间.

References:

[1] Lee S, Levanti K, Kim HS. Network monitoring: Present and future. Computer Networks, 2014,65(2):84–98. [doi: 10.1016/j.comnet.2014.03.007]

[2] Huang G, Chang CW, Chuah CN, Lin B. Measurement-Aware monitor placement and routing: A joint optimization approach for network-wide measurements. IEEE Trans. on Network and Service Management, 2012,9(1):48–59. [doi: 10.1109/TNSM.2012.010912.110128]

[3] Cai ZP. Network measurement technologies, models and algorithms based on active and passive measurement [Ph.D. Thesis]. Changsha: National University of Defense Technology, 2005 (in Chinese with English abstract).

[4] Chen SM. Research on some key issues for internet measured management [Ph.D. Thesis]. Chengdu: University of Electronic Science and Technology, 2009 (in Chinese with English abstract).

[5] NIMI project. http://www.isoc.org/inet2000/cdproceedings/1d/1d_1.htm

[6] Surveyor project. http://www.isoc.org/inet99/proceedings/4h/4h_2.htm

[7] CAIDA project. http://www.caida.org/home/

[8] MOBICOM. In: Proc. of the ACM Int'l Conf. on Mobile Computing and Networking. http://dblp.uni-trier.de/db/conf/mobicom/

[9] SIGCOMM. In: Proc. of the ACM Int'l Conf. on the Applications, Technologies, Architectures, and Protocols for Computer Communication. http://dblp.uni-trier.de/db/conf/sigcomm/index.html

© 中国科学院软件研究所 http://www.jos.org.cn

[10] INFOCOM. In: Proc. of the IEEE Int'l Conf. on Computer Communications. http://dblp.uni-trier.de/db/conf/infocom/

[11] TON. IEEE/ACM Trans. on Networking IEEE. http://dblp.uni-trier.de/db/journals/ton/

[12] JSAC. IEEE Journal of Selected Areas in Communications. http://dblp.uni-trier.de/db/journals/jsac/

[13] TMC. IEEE Trans. on Mobile Computing. http://dblp.uni-trier.de/db/journals/tmc/

[14] Cui Y, Lai ZQ, Wang X, Dai NW, Miao C. QuickSync: Improving synchronization efficiency for mobile cloud storage services. In: Proc. of the Annual Int'l Conf. on Mobile Computing and Networking (MobiCom). Paris: ACM Press, 2015. 592−603. [doi: 10.1145/2789168.2790094]

[15] Guo CX, Yuan LH, Xiang D, Dang YN, Huang R, Maltz D, Liu ZY, Wang V, Pang B, Chen H, Lin ZW, Kurien V. Pingmesh: A large-scale system for data center network latency measurement and analysis. In: Proc. of the 2015 ACM Conf. on Special Interest Group on Data Communication (SIGCOMM). London: ACM Press, 2015. 139−152. [doi: 10.1145/2829988.2787496]

[16] Dong W, Liu YH, He Y, Zhu T. Measurement and analysis on the packet delivery performance in a large scale sensor network. In: Proc. of the 32nd IEEE Conf. on Computer Communications (INFOCOM). Turin: IEEE Press, 2013. 2679−2687. [doi: 10.1109/INFCOM.2013.6567076]

[17] Wang JL, Wei D, Cao ZC, Liu YH. On the delay performance in a large-scale wireless sensor network: Measurement, analysis, and implications. IEEE/ACM Trans. on Networking, 2015,23(1):186−197. [doi: 10.1109/TNET.2013.2296331]

[18] Zhang WX, Chen Y, Yang Y, Wang X, Zhang Y, Hong X, Mao G. Multi-Hop connectivity probability in infrastructure-based vehicular networks. IEEE Journal on Selected Areas in Communications, 2012,30(4):740−747. [doi: 10.1109/JSAC.2012.120508]

[19] Peng SL, Xing, GL, Li SS, Jia WJ, Peng YX. Fast release/capture sampling in large-scale sensor networks. IEEE Trans. on Mobile Computing, 2012,11(8):1274−1286. [doi: 10.1109/TMC.2011.152]

[20] Ji QJ, Dong YJ. A survey on modeling IP network performance characteristics. Journal of China Institute of Communications, 2003, 25(3):151−160 (in Chinese with English abstract).

[21] Zhou H, Li D, Wang YJ. Fundamental problems with available bandwidth measurement systems. Ruan Jian Xue Bao/Journal of Software, 2008,19(5):1234−1255 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/19/1234.htm

[22] Wei AM, Wang HB, Lin Y, Cheng SD. The achievements of bandwidth measurement techniques in IP networks. ACTA Electronica Sinica, 2006,34(7):1301−1310 (in Chinese with English abstract).

[23] Hu ZG, Zhang DL, Zhu AQ, Chen ZW, Zhou HL. SLDRT: A measurement technique for available bandwidth on multi-hop path with bursty cross traffic. Computer Network, 2012,56(14):3247−3260. [doi: 10.1016/j.comnet.2012.06.009]

[24] Downey AB. Using pathchar to estimate internet link characteristics. In: Proc. of the '99 ACM Conf. on Special Interest Group on Data Communication (SIGCOMM). Cambridge: ACM Press, 1999. 241−250. [doi: 10.1145/316188.316228]

[25] Mah BA. Pchar: A tool for measuring internet path characteristics. 2000. http://www.kitchenlab.org/www/bmah/Software/pchar/

[26] Downey A. Clink: A tool for estimating internet link characteristics. 1999. http://allendowney.com/research/clink/

[27] Carter R, Crovella M. Measuring bottleneck link speed in packet-switched networks. Performance Evaluation, 1996,27(8):297−318. [doi: 10.1016/S0166-5316(96)90032-2]

[28] Costantinos D, Parameswaran R, David M. Packet-Dispersion techniques and a capacity estimation methodology. IEEE/ACM Trans. on Network, 2004,12(6):963−977. [doi: 10.1109/TNET.2004.838606]

[29] Kapoor R, Chen L, Lao L, Gerla M, Sanadidi M. CapProbe: A simple and accurate capacity estimation technique. In: Proc. of the 2004 ACM Conf. on Special Interest Group on Data Communication (SIGCOMM). Portland: ACM Press, 2004. 67−78. [doi: 10.1145/1030194.1015476]

[30] Lai K, Baker M. Measuring link bandwidths using a deterministic model of packet delay. In: Proc. of the 2000 ACM Conf. on Special Interest Group on Data Communication (SIGCOMM). Stockholm: ACM Press, 2000. 283−294. [doi: 10.1145/347059.347557]

[31] Jin CK, Lee Y. An end-to-end measurement and monitoring technique for the bottleneck link capacity and its available bandwidth. Computer Networks, 2014,58(1):158−179. [doi: 10.1016/j.comnet.2013.08.028]

[32] Li WW, Zeng B, Zhang DF, Yang JM. Performance evaluation of end-to-end path capacity measurement tools in a controlled environment. In: Proc. of the 3rd Int'l Conf. on Grid and Pervasive Computing (GPC), Vol.5036. Kunming: Springer-Verlag, 2008. 222−231. [doi: 10.1007/978-3-540-68083-3_23]

© 中国科学院软件研究所 http://www.jos.org.cn

[33] Ohkawa N, Nomura Y. Path capacity estimation by passive measurement for the constant monitoring of every network path. In: Proc. of the 16th Asia-Pacific Network Operations and Management Symp. (APNOMS). Hsinchu: IEEE Press, 2014. 1−6. [doi: 10.1109/APNOMS.2014.6996516]

[34] He L, Yu SZ. Methodology and measurement available bandwidth on arbitrary links. Ruan Jian Xue Bao/Journal of Software, 2009, 20(4):997−1013 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3306.htm

[35] Zhou AF, Liu M, Li ZC, Xie GG. Self adaptive method for end-to-end available bandwidth estimation. Journal on Communications, 2008,29(12):37−45 (in Chinese with English abstract).

[36] Huang YC, Lu CS, Wu HK. JitterPath: Probing noise resilient one-way delay jitter-based available bandwidth estimation. IEEE Trans. on Multimedia, 2007,9(4):798−812. [doi: 10.1109/TMM.2007.893343]

[37] Strauss J, Katabi D, Kaashoek F. A measurement study of available bandwidth estimation tools. In: Proc. of the 2003 ACM SIGCOMM Internet Measurement Conf. (IMC). Miami Beach: ACM Press, 2003. 39−44. [doi: 10.1145/948205.948211]

[38] Carter RL, Crovella ME. Measuring bottleneck link speed in packet-switched networks. Performance Evaluation, 1996,27(28): 297−318. [doi: 10.1016/S0166-5316(96)90032-2]

[39] Hu NN, Steenkiste P. Evaluation and characterization of available bandwidth probing techniques. IEEE Journal on Selected Areas in Communications, 2003,21(6):879−894. [doi: 10.1109/JSAC.2003.814505]

[40] Navratil J, Cottrell RL. ABwE: A practical approach to available bandwidth estimation. In: Proc. of the Passive and Active Measurement Workshop (PAM). La Jolla: IEEE Press, 2003. 1−11. http://www.pam2003.org/proceedings.html

[41] Xie Y, Zheng T, Wang YX, Yuan PF. AProbing: Estimating available bandwidth using ACK pair probing. In: Proc. of the 2014 Int'l Conf. on Smart Computing Workshops (MARTCOMP). Hongkong: IEEE Press, 2014. 43−49. [doi: 10.1109/SMARTCOMP-W.2014.7046666]

[42] 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. [doi: 10.1109/TNET.2003.815304]

[43] Ribeiro VJ, Riedi RH, Baraniuk RG, Navrati J, Cottrell L. PathChirp: Efficient available bandwidth estimation for network paths. In: Proc. of the Passive and Active Measurement Workshop (PAM). La Jolla: IEEE Press, 2003. 200−211. [doi: 10.2172/813038]

[44] Melander B, Bjorkman M, Gunningberg P. A new end-to-end probing and analysis method for estimating bandwidth bottlenecks. In: Proc. of the IEEE Global Telecommunications Conf. (GLOBECOM), Vol.27. San Francisco: IEEE Press, 2000. 415−420. [doi: 10.1109/GLOCOM.2000.892039]

[45] Sommers J, Barford P, Willinger W. A proposed framework for calibration of available bandwidth estimation tools. In: Proc. of the Int'l Symp. on Computers and Communications (ISCC). Cagliari: IEEE Press, 2006. 709−718. [doi: 10.1109/ISCC.2006.15]

[46] Dovrolis C, Ramanathan P, Moore D. What do packet dispersion techniques measure. In: Proc. of the 20th IEEE Conf. on Computer Communications (INFOCOM), Vol.2. Anchorage: IEEE Press, 2001. 905−914. [doi: 10.1109/INFCOM.2001.916282]

[47] Liu XC, HE L, YU SZ. Algorithms of accurate network available bandwidth measurement. ACTA Electronica Sinica, 2007.35(1): 68−72 (in Chinese with English abstract).

[48] Wang L, Yang F. Available bandwidth measurement approach with improvements to IGI. Journal of Beijing University of Posts and Telecommunications, 2008,31(5):36−39 (in Chinese with English abstract).

[49] Lao L, Dovrolis C, Sanadidi MY. The probe gap model can underestimate the available bandwidth of multi-hop paths. ACM Sigmetrics Performance Evaluation Review, 2006,36(5):29−34. [doi: 10.1145/1163593.1163599]

[50] Shriram A, Kaur J. Empirical evaluation of techniques for measuring available bandwidth. In: Proc. of the 26th IEEE Conf. on Computer Communications (INFOCOM). Anchorage: IEEE Press, 2007. 2162−2170. [doi: 10.1109/INFCOM.2007.250]

[51] Li M, Wu YL, Chang CR. Available bandwidth estimation for the network paths with multiple tight links and bursty traffic. Journal of Network and Computer Applications, 2013,36(1):353−367. [doi: 10.1016/j.jnca.2012.05.007]

[52] Imai M, Sugizaki Y, Asatani K. A new available bandwidth estimation method using RTT for a bottleneck link. IEICE Trans. on Communications, 2014,97(4):712−720. [doi: 10.1587/transcom.E97.B.712]

[53] Akella A, Seshan S, Shaikh A. An empirical evaluation of wide-area Internet bottlenecks. ACM Sigmetrics Performance Evaluation Review, 2003,31(1):316−317. [doi: 10.1145/781027.781075]

© 中国科学院软件研究所 http://www.jos.org.cn

[54] Zhang DL, Huang WL, Lin C. Locating the tightest link of a network path. ACM Sigmetrics Performance Evaluation Review, 2004,32(1):402−403. [doi: 10.1145/1012888.1005738]

[55] Ribeiro V, Riedi R, Baraniuk R. Spatio-Temporal available bandwidth estimation with STAB. ACM Sigmetrics Performance Evaluation Review, 2004,32(1):394−395. [doi: 10.1145/1012888.1005734]

[56] Zhang DL, Zhang JS, Hu ZG, Zhu XQ. Tight link location based on available bandwidth measurement of subpath. Journal of Computer Applications, 2010,30(12):3141−3144 (in Chinese with English abstract).

[57] Hu N, Li LE, Mao ZM, Steenkiste P, Wang J. Locating Internet bottlenecks: Algorithms, measurements, and implications. In: Proc. of the 2004 ACM Conf. on Special Interest Group on Data Communication (SIGCOMM). Portland: ACM Press, 2004. 41−54. [doi: 10.1145/1030194.1015474]

[58] Lai K, Baker M. Measuring link bandwidths using a deterministic model of packet delay. In: Proc. of the 2000 ACM Conf. on Special Interest Group on Data Communication (SIGCOMM). Stockholm: ACM Press, 2000. 283−294. [doi: 10.1145/347059.347557]

[59] Harfoush K, Bestavros A, Byers J. Measuring bottleneck bandwidth of targeted path segments. In: Proc. of the 22nd IEEE Conf. on Computer Communications (INFOCOM), Vol.3. San Francisco, 2003. 2079−2089. [doi: 10.1109/INFCOM.2003.1209229]

[60] Leland WE, Taqqu MS, Willinger W, Wilson DV. On the self-similarity nature of Ethernet traffic. IEEE/ATM Trans. on Networking, 1994,2(1):1−15. [doi: 10.1109/90.282603]

[61] Liu X, Ravindran K, Loguinov D. A stochastic foundation of available bandwidth estimation: Multi-Hop analysis. IEEE/ACM Trans. on Networks, 2008,16(2):130−143. [doi: 10.1109/TNET.2007.899014]

[62] 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. [doi: 10.1109/TNET.2007.896235]

[63] Almes G, Kalidindi S, Zekauskas M. A one-way packet loss metric for IPPM. RFC 2680, 1999.

[64] Ping. http://www.ping127001.com/pingpage.htm

[65] Adams A, Mahdavi J, Mathis M, Paxson V. Creating a scalable architecture for Internet measurement. In: Proc. of the Inet. Geneva. Switzerland, 1998,14(2):110−114. http://www.isoc.org/inet98/

[66] Shukla S. Sting: A TCP-based network measurement tool. In: Proc. of the 2nd USENIX Symp. on Internet Technologies and Systems (USENIX). Boulder: USENIX Press, 1999,25(5):239−248.

[67] Sommers J, Barford P, Duffield N, Ron A. A geometric approach to improving active packet loss measurement. IEEE/ACM Trans. on Networking, 2008,16(2):307−320. [doi: 10.1109/TNET.2007.900412]

[68] Hu ZG, Zhang DL, Gu LL, Zhang QQ. Embedding method for packet loss self-measurement of video streaming. Ruan Jian Xue Bao/Journal of Software, 2013,24(9):2182−2195 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4338.htm [doi: 10.3724/SP.J.1001.2013.04338]

[69] Paxson V, Almes G, Mahdavi J, Mathis M. Framework for IP performance metrics. RFC2330, 1998.

[70] Sanneck H, Carle G. A framework model for packet loss metrics based on loss runlengths. In: Proc. of the SPIE/ACM SIGMM Multimedia Computing and Networking (MMCN). San Jose, 2000. 177−187. [doi: 10.1117/12.373520]

[71] Han SY, Abu-Ghazaleh NB, Lee D. Efficient and consistent path loss model for mobile network simulation. IEEE/ACM Trans. on Networking, 2016,24(3):1774−1783. [doi: 10.1109/TNET.2015.2431852]

[72] Yajnik M, Moon S, Kurose J, Towsley D. Measurement and modelling of the temporal dependence in packet loss. In: Proc. of the 18th IEEE Conf. on Computer Communications (INFOCOM). New York: IEEE Press, 1999,(1):345−352. [doi: 10.1109/INFCOM.1999.749301]

[73] Yu X, Modestino JW, Tian X. The accuracy of gilbert models in predicting packet-loss statistics for a single-multiplexer network model. In: Proc. of the 24th IEEE Conf. on Computer Communications (INFOCOM). Miami: IEEE Press, 2005. 2602−2612. [doi: 10.1109/INFCOM.2005.1498544]

[74] Hasslinger G, Hohlfeld O. The Gilbert-Elliott model for packet loss in real time services on the Internet. In: Proc. of the 14th GI/ITG Conf. on Measuring, Modelling and Evaluation of Computer and Communication Systems (MMB). Dortmund: IEEE Computer Society Press, 2008. 269−286. http://dblp.org/rec/conf/mmb/2008

© 中国科学院软件研究所 http://www.jos.org.cn

[75] Yu Y, Miller SL. A four-state markov frame error model for the wireless physical layer. In: Proc. of the IEEE Wireless Communications and Networking Conf. Kowloon: IEEE Press, 2007. 2055−2059. [doi: 10.1109/WCNC.2007.385]

[76] Estrada L, Torres D, Toral H. Characterization and modeling of packet loss of a VoIP communication. World Academy of Science, Engineering and Technology, 2010,4(6):926−930.

[77] Ghita D, Argyraki K, Thiran P. Network tomography on correlated links. In: Proc. of the 2010 ACM SIGCOMM Internet Measurement Conf. (IMC). Melbourne: ACM Press, 2010. 225−238. [doi: 10.1145/1879141.1879170]

[78] Ma L, He T, Leung KK, Swami A, Towsley D. Inferring link metrics from end-to-end path measurements: Identifiability and monitor placement. IEEE/ACM Trans. on Networking, 2014,22(4):1351−1368. [doi: 10.1109/TNET.2014.2328668]

[79] Gjoka M, Fragouli C, Sattari P, Markopoulou A. Loss tomography in general topologies with network coding. In: Proc. of the IEEE Global Telecommunications Conf. (GLOBECOM). Washington: IEEE Press, 2007. 381−386. [doi: 10.1109/GLOCOM.2007.78]

[80] Almes G, Kalidindi S, Zekauskas MJ. A one-way delay metric for IP performance metrics. RFC 2679, 1999.

[81] Ishibashi K, Aida M, Kuribayashi SI. Proposal and evaluation of method to estimate packet loss-rate using correlation of packet delay and loss. IEICE Trans. on Information and Systems, 2003,86(11):2371−2379.

[82] Almes G, Kalidindi S, Zekauskas MJ. A round-trip delay metric for IP performance metrics. RFC 2681, 1999.

[83] Demichelis C, Chimento P. IP packet delay variation metric for IP performance metrics. RFC 3393, 2002.

[84] Moon SB, Skelly P, Towsley D. Estimation and removal of clock skew from network delay measurements. In: Proc. of the 18th IEEE Conf. on Computer Communications (INFOCOM). New York: IEEE Press, 1999. 227−234. [doi: 10.1109/INFCOM.1999.749287]

[85] Paxson V. On calibrating measurements of packet transit times. ACM Sigmetrics Performance Evaluation Review, 1998,26(1):11−21. [doi: 10.1145/277858.277865]

[86] Zhang L, Liu Z, Xia CH. Clock synchronization algorithms for network measurements. In: Proc. of the 21st IEEE Conf. on Computer Communications (INFOCOM). New York: IEEE Press, 2002. 50−63. [doi: 10.1109/INFCOM.2002.1019257]

[87] Wang HB, Lin Y, Jin YH, Cheng SD. A new approach for removing clock skew and resets from one-way delay measurement. Acta Electronica Sinica, 2005,33(4):584−589 (in Chinese with English abstract).

[88] Jabbarifar M, Dagenais M, Shameli-Sendi A. Online incremental clock synchronization. Journal of Network and Systems Management, 2015,23:1034−1066. [doi: 10.1007/s10922-014-9331-7]

[89] Li WW, Zhang DF, Xie GG, Yang JM. A high precision approach of network delay measurement based on general PC. Ruan Jian Xue Bao/Journal of Software, 2006,17(2):275−284 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/17/275.htm

[90] Oka Saputra K, Teng WC, Chen TH. Hough transform-based clock skew measurement over network. IEEE Trans. on Instrumentation and Measurement, 2015,64(12):3209−3216. [doi: 10.1109/TIM.2015.2450293]

[91] Zhang QF, Venkatasubramanian VM. Synchrophasor time skew: Formulation, detection and correction. In: Proc. of the 2014 North American Power Symp. (NAPS). Pullman: IEEE Press, 2014. 1−6. [doi: 10.1109/NAPS.2014.6965457]

[92] Georgatos F, Gruber F, Karrenberg D. Providing active measurements as a regular service for ISP'S. In: Proc. of the Passive and Active Measurement Workshop (PAM). Amsterdam: IEEE Press, 2001. http://www.pam2001.org/proceedings.html

[93] Mills DL. Network time protocol (version 3) specification, implementation and analysis. RFC1305, 1992.

[94] Eidson J, Kang L. IEEE standard for a precision clock synchronization protocol for networked measurement and control systems. IEEE Standards, IEEE Std 1588TM-2002, 2008.

[95] Xu Y. Research on the end to end one way delay measurement technology of communication network [MS. Thesis]. Nanjing: Nanjing University of Posts and Telecommunications, 2012 (in Chinese with English abstract).

[96] Chen L, Zhu T, Liu F, Wang W. Modeling and synchronization for IEEE 1588 clock based on Kalman. Lecture Notes in Electrical Engineering, 2014,323:609−618. [doi: 10.1007/978-3-662-44687-4_54]

[97] Mallada E, Meng X, Hack M, Zhang L. Skewless network clock synchronization without discontinuity: Convergence and performance. IEEE/ACM Trans. on Networking, 2015,23(5):1619−1632. [doi: 10.1109/TNET.2014.2345692]

[98] Kim D, Lee J. One-Way delay estimation without clock synchronization. IEICE Electronics Express, 2007,14(23):717−723. [doi: 10.1587/elex.4.717]

© 中国科学院软件研究所 http://www.jos.org.cn

[99] Chen S, Wang J, Qin Y, Zhou X. Accurate and low cost scheme for network path delay measurement. Journal of Software, 2013, 8(6):1398−1404. [doi: 10.4304/jsw.8.6.1398-1404]

[100] Gurewitz O, Cidon I, Sidi M. One-Way delay estimation using network-wide measurements. IEEE Trans. on Information Theory, 2006,52(6):2710−2724. [doi: 10.1109/TIT.2006.874414]

[101] Mota-García E, Hasimoto-Beltran R. A new model-based clock-offset approximation over IP networks. Computer Communications, 2014,53:26−36. [doi: 10.1016/j.comcom.2014.07.006]

[102] Kim H, Ma X, Hamilton BR. Tracking low-precision clocks with time-varying drifts using Kalman filtering. IEEE/ACM Trans. on Networking, 2012,20(1):257−270. [doi: 10.1109/TNET.2011.2158656]

[103] Liu J. An efficient method of estimating the ratio of clock frequency. In: Proc. of the IEEE Military Communications Conf. (MILCOM). Baltimore: IEEE Computer Society Press, 2014. 849−858. [doi: 10.1109/MILCOM.2014.147]

[104] Wang H, Lee KS, Li E. Timing is everything: Accurate, minimum overhead, available bandwidth estimation in high-speed wired networks. In: Proc. of the IMC 2014. Vancouver: ACM Press, 2014. 407−420. [doi: 10.1145/2663716.2663746]

[105] Zhang Y, Duffield N, Paxson V, Shenker S. On the constancy of internet path properties. In: Proc. of the 1st ACM SIGCOMM Internet Measurement Workshop (IMW). San Francisco: ACM Press, 2001. 197−211. [doi: 10.1145/505202.505228]

[106] Alexander G, Crandall JR. Off-Path round trip time measurement via TCP/IP side channels. In: Proc. of the 34th IEEE Conf. on Computer Communications (INFOCOM). Kowloon: IEEE Press, 2015. 1589−1597. [doi: 10.1109/INFOCOM.2015.7218538]

[107] Chan EWW, Chen A, Luo XP, Mok RKP, Li WC, Chang RKC. TRIO: Measuring asymmetric capacity with three minimum round-trip times. In: Proc. of the 7th Conf. on Emerging Networking Experiments and Technologies (CoNEXT). Tokyo: ACM Press, 2011. 1−12. [doi: 10.1145/2079296.2079311]

[108] Padmanabhan VN, Subramanian L. An investigation of geographic mapping techniques for internet hosts. ACM Sigcomm Computer Communication Review, 2001,31(4):173−185. [doi: 10.1145/964723.383073]

[109] Chan EWW, Luo X, Chang RKC. A minimum-delay-difference method for mitigating cross-traffic impact on capacity measurement. In: Proc. of the 2009 ACM Conf. on Emerging Networking Experiments and Technologies (CoNEXT). Rome: ACM Press, 2009. 205−216. [doi: 10.1145/1658939.1658963]

[110] Guerrero CD, Salcedo D, Lamos H. A clustering approach to reduce the available bandwidth estimation error. IEEE Latin America Trans., 2013,11(3):927−932. [doi: 10.1109/TLA.2013.6568835]

[111] Cai ZP, Liu F, Zhao WT, Liu XH, Yin JP. Deploying models and optimization algorithms of network measurement. Ruan Jian Xue Bao/Journal of Software, 2008,19(2):419−431 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/19/419.htm

[112] Wang J, Wang BQ, Zhang XH. Network measurement task deploying algorithm based on reconfiguration model. Journal of Electronics & Information Technology, 2015,37(7):1598−1605 (in Chinese with English abstract). [doi: 10.11999/JEIT141336]

[113] Bouten N, Schmidt RDO, Famaey J, Latré S. QoE-Driven in-network optimization for adaptive video streaming based on packet sampling measurements. Computer Networks, 2015,81:96−115. [doi: 10.1016/j.comnet.2015.02.007]

[114] Lozano F, Gómez G, Aguayo-Torres MC, Cárdenas C, Plaza A. Network performance testing system integrating models for automatic QoE evaluation of popular services: YouTube and Facebook. Wireless Personal Communications, 2015,81(4):1377−1397. [doi: 10.1007/s11277-015-2479-y]

[115] Argon O, Shavitt Y, Weinsberg U. Inferring the periodicity in large-scale Internet measurements. In: Proc. of the 32nd IEEE Conf. on Computer Communications (INFOCOM). Turin: IEEE Press, 2013. 1672−1680. [doi: 10.1109/INFCOM.2013.6566964]

[116] Fontugne R, Mazel J, Fukuda K. An empirical mixture model for large-scale RTT. In: Proc. of the 34th IEEE Conf. on Computer Communications (INFOCOM). Kowloon: IEEE Press, 2015. 2470−2478. [doi: 10.1109/INFOCOM.2015.7218636]

[117] Yin H, Qiao B. Big Data-Driven network information plane. Chinese Journal of Computers, 2016,39(1):126−139 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2016.00126]

[118] Biswas S, Bicket J, Wong E, Musaloiu-ER, Bhartia A, Aguayo D. Large-Scale measurements of wireless network behavior. ACM Sigcomm Computer Communication Review, 2015,45(5):153−165. [doi: 10.1145/2785956.2787489]

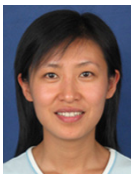
附中文参考文献:

[3] 蔡志平.基于主动和被动测量的网络测量技术、模型和算法研究[博士学位论文].长沙:国防科学技术大学,2005.

© 中国科学院软件研究所 http://www.jos.org.cn

[4] 陈松.互联网测量管理若干关键技术研究[博士学位论文].成都:电子科技大学,2009.
[20] 纪其进,董育宁.IP 网络性能特征模型分析.通信学报,2003,25(3):151−160.
[21] 周辉,李丹,王永吉.可用带宽度量系统中的若干基本问题.软件学报,2008,19(5):1234−1255. http://www.jos.org.cn/1000-9825/19/1234.htm
[22] 韦安明,王洪波,林宇,程时端.IP 网带宽测量技术研究与进展.电子学报,2006,34(7):1301−1310.
[34] 何莉,余顺争.一种测量任意链路可用带宽的方法.软件学报,2009,20(4):997−1013. http://www.jos.org.cn/1000-9825/3306.htm
[35] 周安福,刘敏,李忠诚,谢高岗.自适应的端到端可用带宽测量方法.通信学报,2008,29(12):37−45.
[47] 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.
[48] 王雷,杨帆.改进 IGI 的可用带宽测量方法.北京邮电大学学报,2008,31(5):36−39.
[56] 张大陆,张俊生,胡治国,朱小庆.基于子路径可用带宽测量的紧链路定位方法.计算机应用,2010,30(12):3141−3144.
[68] 胡治国,张大陆,谷丽丽,张起强,陈志伟,周华磊,曹孝晶.一种嵌入视频流的丢包自测量方法.软件学报,2013,24(9):2182−2195. http://www.jos.org.cn/1000-9825/4338.htm [doi: 10.3724/SP.J.1001.2013.04338]
[87] 王洪波,林宇,金跃辉,程时端.一个消除单向时延测量中时钟频差和时钟重置的新方法.电子学报,2005,33(4):584−589.
[89] 黎文伟,张大方,谢高岗,杨金民.基于通用 PC 架构的高精度网络时延测量方法.软件学报,2006,17(2):275−284. http://www.jos.org.cn/1000-9825/17/275.htm
[95] 徐勇.通信网端到端单向时延测量技术的研究[硕士学位论文].南京:南京邮电大学,2012.
[111] 蔡志平,刘芳,赵文涛,刘湘辉,殷建平.网络测量部署模型及其优化算法.软件学报,2008,19(2):419−431. http://www.jos.org.cn/1000-9825/19/419.htm
[112] 王晶,汪斌强,张校辉.基于可重构测量模型的网络测量任务部署算法.电子与信息学报,2015,37(7):1598−1605. [doi: 10.11999/JEIT141336]
[117] 尹浩,乔波.大数据驱动的网络信息平面.计算机学报,2016,39(1):126−139. [doi: 10.11897/SP.J.1016.2016.00126]

胡治国(1977－),男,山西灵石人,博士,讲师,CCF 专业会员,主要研究领域为网络性能测量,音视频质量评价.



关晓蔷(1979－),女,讲师,CCF 专业会员,主要研究领域为机器学习.

田春岐(1975－),男,博士,副教授,主要研究领域为网络性能测量,云计算.

曹峰(1980－),男,博士,讲师,CCF 专业会员,主要研究领域为粗糙集,数据挖掘.

杜亮(1985－),男,博士,讲师,CCF 专业会员,主要研究领域为数据挖掘.

© 中国科学院软件研究所 http://www.jos.org.cn