

可用带宽与瓶颈定位集成测量工具 iPathneck^①

朱 丽, 黎文伟

(湖南大学 信息科学与工程学院, 长沙 410082)

摘要: 目前大部分的网络测量工具都是采用主动测量方法, 但是由于大量的探测包注入会给网络带来流量负载并影响测量精度. 如何在保证测量精度的前提下减小测量开销, 成为值得关注的问题. 提出的集成测量工具 iPathneck 在 Pathneck 工具的基础上采用速率模型提高可用带宽的测量精度, 实现了可用带宽和瓶颈定位探测任务的归并, 减小了测量开销并提高测量精度. 实验表明: iPathneck 可用带宽和瓶颈定位的测量准确性和算法时间收敛性都有所提高, 任务归并能有效地减小测量开销.

关键词: 主动测量; 可用带宽; 瓶颈定位; 速率模型; 测量开销

Combined Tool for Measuring of Available Bandwidth and Locating Internet Bottleneck-iPathneck

ZHU Li, LI Wen-Wei

(College of Information Science and Engineering, Hunan University, Changsha 410082, China)

Abstract: Presently, most of the network measurement tools are active. But they bring heavy overhead and impact the measurement accuracy. We concern how to ensure measurement accuracy with reducing the overhead. In this work, a new measurement scheme iPathneck based on Pathneck is proposed to perform better than Pathneck on measuring ABW by using probe rate model. iPathneck combines tasks to measure ABW and to locate Internet bottleneck. Experiment will show that iPathneck achieve shorter convergence time, lighter measurement overhead and higher accuracy.

Key words: active probing; available bandwidth; locating Internet bottleneck; probe rate model; measurement overhead

准确的网络性能指标测量是进行网络流量工程的依据, 确保我们充分有效的利用网络资源. 端到端路径可用带宽和瓶颈位置是网络服务提供商 ISP 了解网络状况, 进行网络优化管理的两个重要依据.

可用带宽测量模型可分为探测间距模型^[1,2] (probe gap model, PGM)和探测速率模型^[3,4] (probe rate model, PRM). PGM 模型提取到达目的端的两个连续探测分组之间的时间间隔信息, 分析探测分组发送间隔与到达间隔的变化, 估计可用带宽. PRM 在源端以一定速率向目的端发送探测分组. 当探测速率达到一定程度时, 会造成网络拥塞, 探测分组的排队时延瞬间增大, 使网络路径从不拥塞状态转

入自拥塞临界状态. 对应的自拥塞临界速率, 被认为是端到端路径的可用带宽^[2]. 间距模型存在对背景流敏感不稳定的问题, 速率模型则会在短时间内急剧增加网络的流量负载, 干扰网络的正常运行并影响测量精度.

瓶颈定位工具也普遍存在测量不准确或测量开销大等问题. 在众多的瓶颈定位工具中 Pathneck 是一个高效的, 轻量级测量工具, 能以较少的探测注入以 80% 的准确率定位到瓶颈链路. 但是该工具只能粗略地估计可用带宽的范围. 因此, 在保留 pathneck 瓶颈定位高准确性和低开销优势的情况下改进其可用带宽测量精度, 对减小测量开销有很大的现实意义.

① 基金项目:国家自然科学基金(61173168);湖南省科技计划(2011FJ3236);东莞市高等院校科技计划(2011108101010)

收稿时间:2012-03-15;收到修改稿时间:2012-05-19

1 Pathneck介绍

1.1 Pathneck 的瓶颈定位原理

Pathneck 采用递归探测包列(Recursive Packets Train- RPT)^[5]定位瓶颈. RPT 由负荷分组(load packet)和测量分组(measurement packet)构成, 其中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 RPT 到达该跳路由时的长度. 发送一系列 RPT, 筛选出 RPT 长度达到最大的转折点链路, 作为候选瓶颈链路. 多次发送 RPT, 最后根据置信度, 统计结果, 确定瓶颈链路瓶位置.

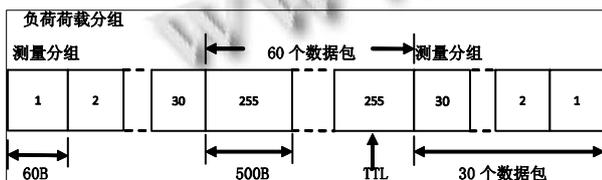


图 1 递归探测包列(RPT)

图 1 为 RPT 递归探测包列结构. 其中格子代表 UDP 数据包, 格子中的数值代表生命周期值 TTL. 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增减. 测量分组是标准的 traceroute 数据包, 即 60 字节的 UDP 数据包.

1.2 Pathneck 的可用带宽测量原理

Pathneck 可用带宽的计算类似 cprobe 工具^[6], 假设瓶颈链路出现在窄链路^[2]上. 计算公式 $R=ts/g$, R 为 RPT 递归探测包列到达目的端的速率, ts 为 RPT 的总长, g 为 RPT 经过每跳路由后产生的两个 ICMP 分组间距的平均值, 粗略地代替窄链路可用带宽容量.

1.3 Pathneck 可用带宽测量不准确原因分析

首先: Pathneck 假设瓶颈链路是窄链路^[5], 但这种假设常常不成立. 瓶颈链路分为: 容量瓶颈和负载瓶颈. 容量瓶颈^[1]可用带宽小是受限于小的带宽容量. 负载瓶颈^[5]可用带宽小是由于该链路虽带宽容量大, 但处于流量交叉汇集点, 流量负载非常大. 负载瓶颈背景流更复杂, 更不稳定. 在负载瓶颈情况下,

Pathneck 可用带宽测量偏差会很大.

其次: 网路测量工具^[1,2]证明能用探测包列探测端到端路径可用带宽. 但是必须控制好探测包间初始间距, 而递归探测包列采用背靠背发送, 不能很好控制探测包间距.

最后: 因为不能控制 RPT 经过每跳路由时的结构, RPT 结构变化会影响可用带宽的测量准确性.

2 iPathneck的改进设计

2.1 iPathneck 设计思想

iPathneck 沿用 Pathneck 的瓶颈定位原理, 采用 RPT 递归包列定位瓶颈, 但依据速率模型测量端到端路径可用带宽. 即在源端以不同的速率发送探测包, 搜索路径由非拥塞状态转入自拥塞状态的临界速率, 以此作为端到端路径可用带宽. 临界速率采用三元搜索算法进行搜索: 交替发送 RPT 递归探测包列和辅助探测包. 如图 2 所示, 以大于可用带宽的速率开始发送一系列递归探测包列, 然后逐渐减小发送速率发送; 以小于可用带宽的速率开始发送一系列辅助探测包列, 然后逐渐加大速率发送; 两种不同结构探测包分别从可用带宽范围的上下界逐渐逼近路径自拥塞临界速率. 在发送过程中根据探测包测量分组间距变化情况和在源端的发送速率动态的改变可用带宽的上下界值预测值, 并作为参数确定探测包的发送速率.

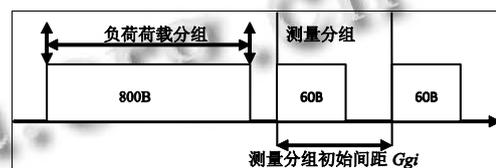


图 2 辅助探测包结构

2.2 iPathneck 引入辅助探测包

RPT 根据探测包列的长度改变值确定瓶颈位置, 其发送速率小于或等于可用带宽时, 网络路径不会发生拥塞, 此时不能确定瓶位置. 因此, 在搜索路径拥塞状态临界速率的过程中, 在小于可用带宽真实值范围内, 可以采用其他负载较小的探测包代替 RPT 进行搜索. 这样既不影响瓶颈定位, 也能充分利用 RPT 的速率信息获取测量可用带宽. 并能减小网络负载.

2.3 iPathneck 辅助探测包结构及自拥塞临界速率的判定

图 2 所示为辅助探测包结构^[2], 由一个 800B 和两个

60B 的 UDP 数据包构成。800B 的数据包为负荷荷载，两个 60B 的数据包为测量分组。800B 设置适当的 TTL 使之在目的端被丢弃，以保证两个测量分组在同一队列区域。两个测量分组在目的端则因为目的端口不可达而被丢弃，产生两个 ICMP 应答分组。这两个应答分组返回源端的间距被认为是测量分组到达目的端的时间间距。在源端计算测量分组初始时间间距与对应 ICMP 应答分组时间间距的差值和测量分组差值变化率。

采用测量分组间距变化率来判定是否达到自拥塞临界速率。辅助探测包采用仅有的这个测量分组间距变化率判定网络路径是否发生了拥塞。递归探测包列采用一组中最大的测量分组间距变化率判定网络路径是否发生了拥塞。

2.4 iPathneck 三元算法发包流程

交替发送辅助探测包和递归探测包列，当发送速率等于临界速率或探测包测量分组间距变化率小于某一阈值 δ 时，搜索结束。下面为三元线性搜索算法动态确定发包速率的计算公式，如下所示：

$$R_{min} \text{ Mbps}, R_{max} \text{ Mbps}$$

$$R_g = R_{min} + (R_{max} - R_{min})/3$$

$$R_h = R_{min} + 2*(R_{max} - R_{min})/3$$

$$G_g = (G_{g0} - G_{gi})/G_{gi}$$

$$G_h = (G_{h0} - G_{hi})/G_{hi}$$

R_{min} Mbps, R_{max} Mbps 分别为可用带宽预测范围上下界值， R_g 、 G_g 、 G_{gi} 、 G_{g0} 分别为辅助探测包的发送速率，测量分组的时间间距变化率，测量分组在源端的初始间距以及对应的 ICMP 应答返回到源端时的时间间距。

R_h 、 G_h 、 G_{hi} 、 G_{h0} 则分别为递归探测包列的发送速率，其在每跳路由处首尾测量分组的间距变化率，首尾测量分组在源端的初始间距，以及对应的两个 ICMP 应答包的间距。 G_{hmax} 为发送一次递归探测包列得到的一组测量分组间距变化率 $\{G_{h1}, G_{h2} \dots G_{hn}\}$ 中的最大值。

当 G_g 、 G_{hmax} 的大小在 1%-1.5% 之间或者 R_g 、 R_h 之间的差值小于或等于 1.5Mb/s 时，结束搜索，得到一个网络端到端路径可用带宽测量值。

3 实验

3.1 实验目的

试验需要网络拓扑和流量信息，但是现实中很难

获取。所以，除了在真实网络环境中对 iPathneck 与 Pathneck 进行可用带宽测量和瓶颈定位结果比对外，将进行模拟实验，精确控制实验环境，测试改进工具各个方面的性能。

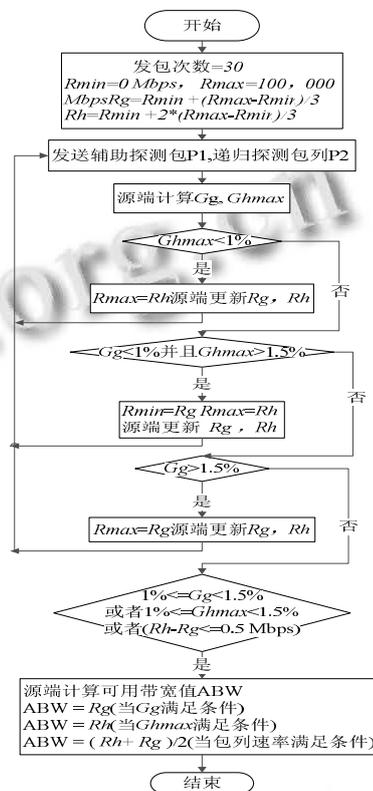


图 3 iPathneck 三元算法发包流程图

3.2 真实网络实验

在普通 PC 机上运行 iPathneck 和 Pathneck，以不同的 IP 地址为目的端，进行可用带宽测量和瓶颈定位。PC 机 CPU 配置为 2.4GHZ，内存容量为 1GB，操作系统 ubuntu10.0 源端 IP 地址均为 192.168.128.131。

表 1 实验结果汇总

工具编号	目的端 IP	可用带宽	瓶颈链路位置
pathneck	81.161.208.1	4.389	192.168.128.2
ipathneck	81.161.208.1	3.408	192.168.128.2
pathneck	72.30.38.140	4.273	192.168.128.2
ipathneck	72.30.38.140	2.576	192.168.128.2
pathneck	74.125.71.160	4.207	192.168.128.2
ipathneck	74.125.71.160	2.917	192.168.128.2
pathneck	106.11.24.2	4.336	192.168.128.2
ipathneck	106.11.24.2	2.619	192.168.128.2

从表 1 可以看出，iPathneck 和 Pathneck 测得各路

径的瓶颈位置均 192.168.128.2. 说明该链路为个路径的必经之路, 且瓶颈位置比较稳定. 两工具测得的结果相同. 在可用带宽测量方面 iPathneck 测量值比 Pathneck 估计上界值要低, 并且两工具的测量值均比较稳定.

3.3 模拟实验

3.3.1 实验环境

图 4 为实验网络拓扑图: 采用 ns2 模拟实验, 精确控制网络上的竞争流量. iPathneck 是端到端路径的网络测量工具, 因此采用线性拓扑. 普通链路容量配置依据 Dummynet^[1], 链路 X,Y 的容量, 以及各链路的背景流量根据实验时设计的场景配置. 背景流量配置采用工具 light-trace, heavy-trace 和 UDP 数据包产生器.

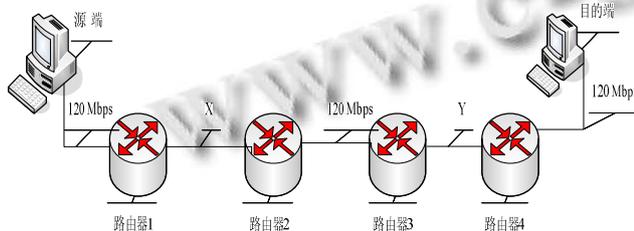


图 4 实验网络拓扑图

3.3.2 实验结果及分析

1) iPathneck, Pathneck 瓶颈定位准确性比较

在不同的容量瓶颈和负载瓶颈情况下运行 iPathneck, Pathneck 各 20 次, 再统计分析测量结果.

图 5 表明: Pathneck 准确率可达到 80%, iPathneck 准确率可达到 90%. iPathneck 较 Pathneck 工具瓶颈定位能力有所改进. iPathneck 可用带宽测量准确性改善可能是提高瓶颈判断准确性的主要原因.

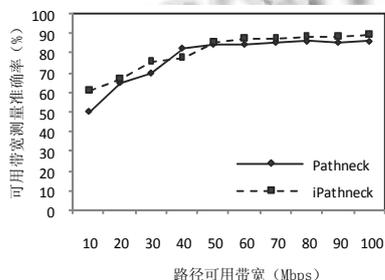


图 5 瓶颈定位准确性比较

2) 容量瓶颈下 iPathneck, Pathneck 可用带宽测量比较

在容量瓶颈下进行, 运行 iPathneck, Pathneck 各 20 次. 再统计分析测量结果.

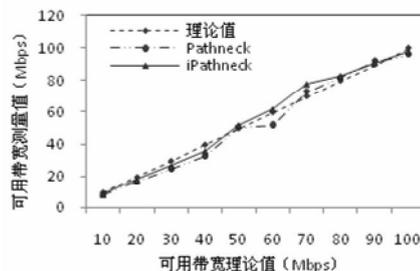


图 6 容量瓶颈下可用带宽测量准确性比较

图 6 表明: iPathneck 和 Pathneck 的可用带宽测量准确性都很高, 能达到 90% 以上. 容量瓶颈是最简单情况, 链路容量是可用带宽唯一的决定因素, 且背景流量很小, 不会影响可用带宽测量.

3) 负载瓶颈下 iPathneck, Pathneck 可用带宽测量比较

在负载瓶颈下, 运行 iPathneck, Pathneck 各 20 次. 再统计分析测量结果.

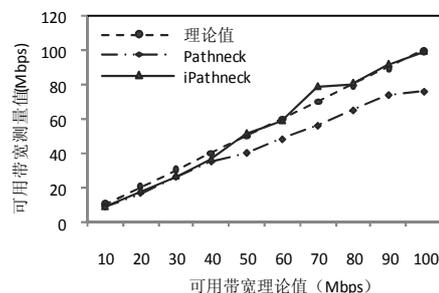


图 7 负载瓶颈下可用带宽测量准确性比较

图 7 表明, 在负载瓶颈下, Pathneck 可用带宽测量值较真实值偏低. iPathneck 可用带宽测量值跟真实值非常接近, 准确率达到 89% 以上. 差别成因: Pathneck 测量机制不完善, 只能粗略的估计上下界值; 而 iPathneck 采用速率模型可以准确的测量可用带宽.

4) iPathneck, Pathneck 测量时间收敛性比较

根据不同场景测量实验统计这两个工具的收敛时间.

图 8 表明, 总体上 iPathneck 较 Pathneck, 收敛速度显著加快. 主要原因: 三元搜索算法动态的缩小可用带宽预测范围, 大大缩短了测量时间.

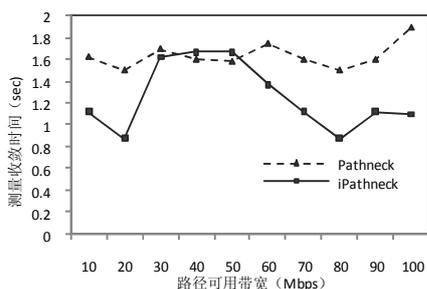


图8 算法时间收敛性比较

4 结语

iPathneck 融合了递归探测包列模型和速率模型, 集成了可用带宽测量和瓶颈定位功能。采用速率模型提高了可用带宽的测量精度, 并提高了瓶颈位置准确的准确性。引入辅助探测包协助递归探测包列采用三元搜索算法大大缩短了测量时间。该工具通过任务归并, 减小测量开销, 达到了测量精度与测量开销之间的平衡。

但是 iPathneck 对于与瓶颈链路可用带宽相近的链路仍然无法识别, 下一步工作将致力于改进这个缺陷。

参考文献

- HU N, Steenkiste P. Evaluation and characterization of available bandwidth probing techniques. *IEEE JSAC*, 2003, 21(6): 879–864.
- Salehin KM, Rojas-Cessa R. Schemes to Measure Available Bandwidth and Link Capacity with Ternary Search and Compound Probe for Packet Networks. *IET Commun*, 2010, 4(2):240–252.
- Antoniades D, Athanatos M, Papadogiannakis A, et al. Available bandwidth measurement as simple as running wget. *Proc PAM*, Adelaide, Australia, 2006.
- Sommers J, Barford P, Willinger W. Laboratory-based calibration of available bandwidth estimation tools. *Micro-processors and Microsystems*, 2007,31(4).
- Hu NN, Li ZQ, Mao M. Locating Internet Bottlenecks: Algorithms, Measurements, and Implications. *SIGCOMM 04*, Aug. 30-Sept.3, 2004,Portland, Oregon, USA.
- Goldenberg D, Qiu L, Xie H, Yang YR, Zhang Y. Optimizing Cost and Performance for Multihoming. *Proc. ACM SIGCOMM*, August 2004.

(上接第 105 页)

ication system. Gateway to the 21st Century 2nd International Conference on Personal Communications. Ottawa, Ont. 12-15 Oct.1993:975–979.

- Wong CY, Cheng RS, Ietaief KB, et al. Multicarrier of DM with adaptive subcarrier, bit, and power Allocation. *IEEE J. Select. Areas Commun*, 1999,17(10):1747–1758.
- Kim I, Lee HL, Kim B, Lee YH. On the use of Linear programming for dynamic subchannel and bit allocation in multiuser OFDM. *IEEE Globecom*,2001,6:3648–3652.
- Yin H, Liu H. An efficient multiuser loading algorithm for OFDM based broadband wireless systems. *IEEE Globecom*,

2000,1:103–107.

- Shen ZK, Andrews JG, Brian L, Evans S. Adaptive Resource Allocation in Multiuser OFDM Systems With Proportional Rate Constraints. *IEEE Trans. on Wireless*, 2005,(4):2726–2737.
- Mao ZW, Wang XM. Efficient Optimal and Suboptimal Radio Resource Allocation in OFDMA System. *IEEE Trans. on Wireless Communications*. 2008,7(2): 440–445.
- 马莉. MATLAB 语言实用教程.北京:清华大学出版社, 2010. 141–176.

(上接第 122 页)

参考文献

- 柴若楠,曾文献. 音视频同步技术综述. *计算机系统应用*, 2011,20:223–226.
- Kuo CC, Chen MS, Chen JC. An adaptive transmission scheme for audio and video synchronization based on real-time transport protocol. 2001 IEEE Int. Conference on Multimedia and Expo(ICME2001), Tokyo, 22-25 Aug. 2001:403–406.

- Palacharla S, Karmouch A, Mahmoud SA. Design and Implementation of a Real-time Multimedia Presentation System using RTP. *Computer Software and Applications Conference, (COMPSAC'97)*. 1997: 376–381.
- 董春兵. 音视频同步的研究与实现[硕士学位论文]. 长春: 吉林大学, 2007.
- 潘鹏, 杜旭, 叶婷. RTP/RTCP 实时传输协议的研究与 Linux 实现. *计算机工程与应用*, 2005,41(24):105–107.