一种基于 Netflow 特定流记录的往返时延估计方法

张晓宇[1,2+]， 龚 俭[1,2+]， 吴 桦[1,2]

[1](东南大学 计算机科学与工程学院,江苏 南京 210096)
[2](江苏省计算机网络技术重点实验室,江苏 南京 210096)

摘要：时延对于网络管理有着重要的意义，测量时延主要采用主动测量和被动测量这两种方法。主动测量需要主动发送报文，而被动测量需要在多个测量点通过捕获在网络信道中传输的报文信息协同来实现，并需要时钟同步。本文首次采用单点被动测量的方法，根据Netflow中特定的流记录，并结合统计的处理方式来估计网络之间的往返时延。通过实验表明，该方法可以基于现有的路由器提供的Netflow数据来测量网络延迟，且具有较小的误差。
关键词：被动测量，Netflow，往返时延

A method of estimating the round-trip time based on the specifically flow records in Netflow

Zhang xiao-yu　　Gong jian　　Wu hua

(School of Computer Science and Engineering，Southeast Univ．，Nanjing 210096)

(Jiangsu Province Key Laboratory of Co mputer Networking Technology，Nanjing 210096)

Abstract：Latency is significant to the network management. Generally，latency can be measured actively or passively. Active measurement needs to send packets actively, and the passive approach requests capturing packets transmitted in network link at more than one measurement points cooperatively, and synchronized clocks. In this paper, combined with statistical process, a new round-trip latency measurement method is proposed that based on single passive measurement point and some specific flow records in Netflow data. Test results show that this method can use the Netflow data from current COTS router to estimate round-trip latency with acceptable error.
Keywords：passive measurement, Netflow; round-trip time

1　引言

在网络管理中，网络时延有非常重要的指标作用，它是计算机网络的重要性能测度，对提供QoS保证、网络规划、差错和拥塞检测、网络性能测试也有重要的作用。

网络时延分为单向时延和往返时延。单向时延表示从发送端发送数据开始到接收端收到该数据所经历的时间；往返时延RTT(Round-Trip Time)表示从发送端发送数据开始，到发送端收到来自接收端的确认，总共经历的时间[1],[3]。

目前，测量网络时延主要采用主动测量和被动测量这两种方式。通常意义上的网络时延测量是指：测量报文往返的时间。主动测量通过发送测量报文来获取链路的延迟，一般通过ping或tracert命令来实现；被动测量不发送测量报文，而是根据具体情况选取正在通信的报文作为测量报文，被动地实现网络时延的测量，具体的实现可以通过两点监控或者单点监控来完成。两点监控测量方法要求所有的设备必须时间同步,而单点监控利用了TCP的确认机制，同时需要记录每个报文的类型、到达时间等详细信息[5]。

Cisco 公司的 Darren Kerr 和 Barry Bruins 在1996 年开发的一套网络流量监测技术——Netflow，目前已内建在大部分

作者简介：
张晓宇，1982年，男， 硕士研究生，xyzhang@njnet.edu.cn；
龚 俭， 男，1957年，教授，博士生导师，CCF高级会员，jgong@njnet.edu.cn；
吴 桦， 女，讲师，博士研究生， hwu@njnet.edu.cn ..

Cisco 路由器和交换机上，Juniper、Ex-treme等网络设备供货商也支持Netflow技术，它已逐渐成为事实的标准[8]。

由于Netflow技术的广泛应用，通过分析日常Netflow采集到的统计数据，利用其中的信息来进行时延的相关研究对于网络管理来说具有重要的意义，可以大致掌握网络的性能状况以及全网通信的正常延迟基线，并以此为依据为日后可能出现的通信异常进行评估和管理。

Netflow在组流时对报文进行了抽样，且在流记录中仅仅记下了首尾报文的到达时间（这里指流记录中与时间相关的测度），这样使得传统的被动测量中的单点监控方法无法直接进行运用。

本文首次根据 Netflow 数据中特定的流记录，利用其中的一些信息，采用了统计的处理方式，提出了一种估计往返时延的方法。该方法本质上属于被动测量，因此不产生流量，并不影响网络的正常通信，在一定的误差范围之内，具有很好的效果。

本文余下部分的组织结构为：第 2 节描述了 Netflow 流记录的相关信息，第 3 节提出了具体的估计往返时延的算法，第 4 节给出实验的设计与结果，第 5 节对文章的工作进行了总结，并指出下一步的研究工作。

2 Netflow 的流记录

在Netflow中，流被定义为两端点间单一方向连续的数据流，这意味着每一个网络的连接都会被分别记录成两个流数据，使用源、宿IP，源、宿端口号，协议类型，服务类型以及路由器口输入接口字段来区分每一个流[4]。

当一个报文被抽中后，路由器会检视这七个字段来判断这个报文是否属于任何已记录的流，是则将新收集到的报文的相关流量信息整合到对应的流记录中，如果找不到，便产生一个新的流记录来储存相关的流量信息[6]。

当下面四种情况任何一种成立，路由器就会通过UDP 数据包将终止的流记录输出到事先指定的Netflow数据收集设备上：当报文内字段flag 显示传输协议中传输完成的信息如TCP FIN或RST时；流量停止超过15秒；流量持续传送，每30 分钟会自动终止；当“流cache”无空间时[7]。

Netflow最初版本为版本1，版本5是获得最广泛使用的版本 。目前Netflow最新版本是版本9[8]。

本文中所提到的Netflow数据格式（V5）部分字段如下：

Start|End|Sif|secIPaddress|srcp|Dif|DstIPaddress|Dstp|P|Fl|Pkts|octets|

流起始时间|流结束时间|路由器输入接口|源 IP 地址|源端口|路由器输出接口|目的 IP 地址|目的端口|协议种类|TCP 控制标记|报文数|字节数|

本文所提出的方法需要利用到的数据包括流起始时间、流结束时间、报文数。

3 基于特定长流的往返时延估

3．1 基本思路

如图 1 所示：设有一个报文（该报文是从 A 发往 B）经过链路中 C 点的时刻为 T_1，经时间 Δt_1 后该报文到达 B,B 在 Δt_B 后给 A 发送一个报文，经 Δt_2 该报文到达 A 之后，

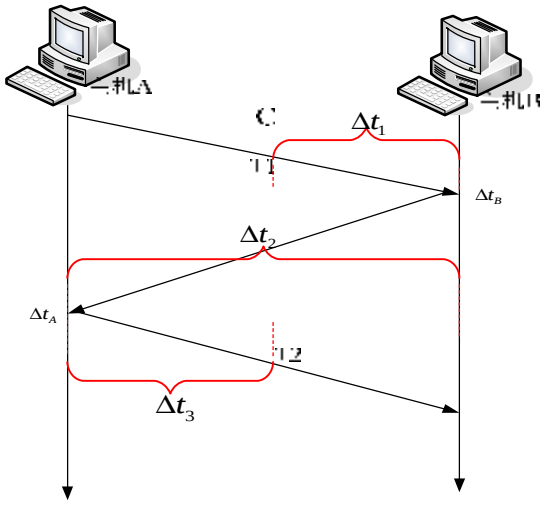

图 1 报文测量模型

A 经过 Δt_A 时间后发送下一个报文，设该报文经过时间 Δt_3 后到达 C 点的时刻为 T_2，

则有：

$$T=T_2-T_1=\Delta t_1+\Delta t_B+\Delta t_2+\Delta t_A+\Delta t_3\cdots\cdots (1)$$

其中T_1和T_2已知，当下列条件之一成立时：

1) Δt_A（或Δt_B）可以忽略不计；

2) Δt_A（或Δt_B）可以估计出来。

则可以求得主机 A,B 间(或 B，A)的往返时延。

当 A 在收到报文后若立即（或经过一个相对于往返时延来说很短的时间）发送下一个报文，则Δt_A可以忽略不计，满足第一个条件，在实际的通信中，可以根据特定的协议特征寻找这样的报文来进行计算，如在一次 TCP 的数据传输中，当主机 A 收到 B 发来的确认报文（即 ack 报文）后，如果还有数据需要传输，将会在一个时间段后继续发送数据，通常该时间可以忽略[2]，B 收到数据后发送 ack 报文，这样可以近似的认为主机 B 对 A 的往返时延为相邻的 ack 报文链路中 C 点的时间之差。

3．2 利用 Netflow 流记录估计 ack 报文数目

上述方法的实质就是求相邻 ack 报文对经过采集点的时间差，同时，这两个相邻的报文对之间需要存在另一个（或一组）反方向的报文，但由于 Netflow 采集器对报文进行了抽样（本文中抽样比为：256：1），无法判断两个 ack 报文是否是相邻的，本文采用了统计的方法进行了处理，设 Netflow 采集路由器采集到的两主机对之间第一个 ack 流的开始时间为 T1（即为首个 ack 报文被采集的时刻）,在此后的一个时间粒度内，记录下每一个采集到的两主机之间的其它 ack 流的起始时间为$T_2.....T_n$（为了保证两主机之间是较长时间的通信，采集到足够多的 ack 报文，本文取 n>5），最后一个流的结束时间为T_n'，假设 ack 流流内报文数目分别为：$p_1, p_2,\cdots\cdots, p_n$，令$\overline{p}=\frac{1}{n}\sum_{i=1}^{n}p_i$，这样得到原始 ack 报文数目为$f(\overline{p})\cdot\sum_{i=1}^{n}p_i$，其中$f(\overline{p})$为经过大量的实验，并且与 Netflow 数据流进行对比，得到经验公式（2）（此公式只是在本文中 256:1 的抽样比下，针对一个时间粒度内的特定流的估计，对于其它数据流可以利用实验的方法类似得到经验公式）[9]。

以上估计出的都是 ack 报文，并不包括非 ack 报文。虽然在通信中可能两主机在互相传输数据，但对于一个单向的数据传输来说，依然符合图 1 中的测量模型，因此相邻 ack 报文的时间差仍可以作为往返时延的估计。

在估计出了原始报文数后，如果在数据传输过程中没有发生超时或重传，则可以很容易的得到往返时延；当发生了超时或重传，此时某些相邻 ack 报文对的时间差已经不能再用来估计，因此，在估计前，需要先进行判断。

$$f(\overline{p})=\begin{cases}160, 1\le\overline{p}<1.6\\ 125, 1.6\le\overline{p}<1.8\\ 65, 1.8\le\overline{p}<2.5\\ 40, 2.5\le\overline{p}\le3.5, \overline{p}\ge10.5\\ 70, 3.5<\overline{p}<4.0\\ 60, 4.0\le\overline{p}<4.5\\ 43, 4.5\le\overline{p}<5.5\\ 60, 5.5\le\overline{p}<6.5\\ 110, 6.5\le\overline{p}<7.5\\ 60, 7.5\le\overline{p}<8.5\\ 50, 8.5\le\overline{p}<10.5\end{cases}\quad\cdots(2)$$

当发生了超时或重传时，ack 报文的发送时间间隔会有很大的变化，则相邻 ack 流之间的起始时间间隔随之会有一定的抖动，具体可以体现在这组间隔值的方差上，当方差很大的时，可以推知某些 ack 报文到达的

间隔时间不均匀，报文到达速度变化较大，当标准方差很小的时候，可以认为主机间的数据传播速度变化不大，此时在数据传输中很少（甚至没有）发生重传或超时，本文通过相邻 ack 流之间的起始时间间隔的均值 $\overline{\Delta t}$ 和方差 S 的阀值来判定在 TCP 传输中是否发生了重传或超时。

3．3 算法的实现

具体的往返时延估计算法如下：

1) 在 Netflow 流记录中找到 ack 流（即该流内的所有报文都为 ack 报文）；

2) 采用统计的方法计算出一个时间粒度内（本文中取为五分钟）两主机对之间 ack 流流内原始报文数；

3) 计算相邻 ack 流之间的起始时间间隔值 Δt_i 及其均值 $\overline{\Delta t}$ 和标准方差 S；

4) **设定 $\overline{\Delta t}$ 和 S 的阀值 a, b，若 $\overline{\Delta t} \le a, S \le b$，则在 TCP 传输数据中，没有发生重传或超时，按如下公式计算往返时延：**

$$T=(T_n'-T_1)/(f(\overline{p})\cdot\sum_{i=1}^{n} p_i-1).....(3)$$

把上述值作为估计值，否则不再进行估计。

在 1)中，具体的方法为找到协议类型为 6（即为 TCP 流）标志位为 2（ack 位置 1）的流，计算该流的流内报文平均字节数（流内字节数/流内报文数），本文通过大量抓包实验表明，当该值小于 60 字节（抓包实验表明 94%的 ack 报文小于 60 字节），可以近似的认为该流内的所有报文都为 ack 报文。

对于 4)，a 取为 15s，b 取 10s，具体值可以通过丢包实验来确定。

4 实验设计及结果

4．1 实验设计

本文提出了当 TCP 数据传输过程中不发生重传或超时时，利用相邻 ack 报文对经过采集点的时间差来估计往返时延的方法，但由于 Netflow 组流时对报文进行了抽样，这样必须确保链路上有大量相邻 ack 报文流可供抽样，同时，考虑到在数据传输中可能存在短暂的超时或重传，这样需要有足够长的通信时间以便恢复正常通信，从而得到正常情况下的报文。为此，本文设计了如下的实验，如图所示：主机 1 分别从主机 2，3，4 长时间（超过 15 分钟）通过 ftp 下载一个大文件，Netflow 采集器采用 UDP 协议来发送输出的数据报文到数据处理器，同时，为了得到较准确的往返时延，主机 1 不断的 ping 主机 2，3，4，把得到的 RTT 作为往返时延的实测值。

在实际实验中，主机 1 为本人的实验机器，主机 2，3，4 为根据主机 1 与其之间的跳数(用 tracert 命令得到)和地理位置的不同而选择的一些高校的 FTP 服务器站点，Netflow 采集器从主机 1 与主机 2，3，4 通信链路中的某点采集数据，且距离主机 1 较近，RTT<1ms。

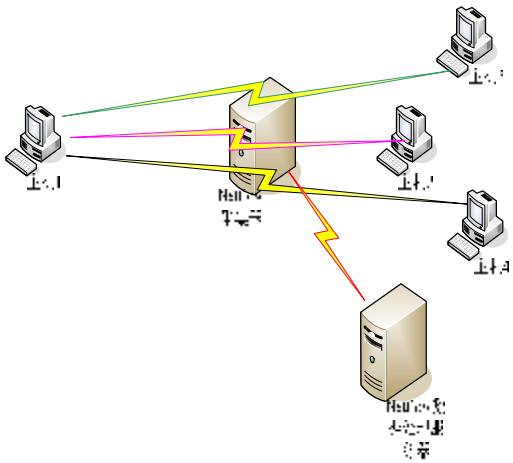

图 2　实验环境拓扑

4．2 实验结果

根据在第 3 部分介绍的方法，进行了多次实验，把估计出的往返时延（称为估计值）与通过 ping 得到的往返时延（称为实测值）进行了比较，如表 1 所示。

表 1 估计值与实测值的比较

FTP 站点	跳数	连接时长（分钟）	实测值（平均值 ms）	估计值（平均值 ms）
山东大学（济南）	14	60	67	60.68
清华大学（北京）	13	60	38	34.51
上海交大（上海）	11	40	18	19.43
河南理工大学（焦作）	16	120	62	57.12

在实际实验中，本文计算出的是每个时间粒度内（5 分钟）往返时延的平均值，为了便于制表，先进行了统计的处理。

此外,定义了估计值相对于实测值的误差百分比：

$$h = \frac{|estRT - RT|}{RT} \times 100\% \quad \text{..... (4)}$$

其中，estRT 为估计值，RT 为实测值，图 3 是估计值相对于实测值的误差分布图。

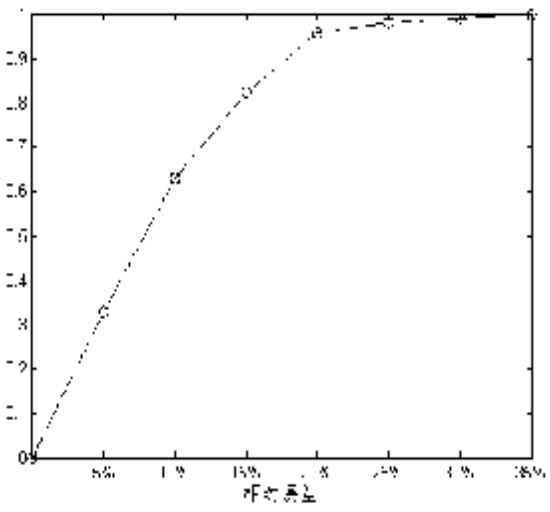

图3 相对误差分布曲线

由图 3 可知，本文所提出的方法对于往返时延的估计具有很好的效果，超过 90%的估计值的相对误差百分比能控制在 20%以下。

5 结束语

本文提出了一种利用 TCP 正常数据传输中相邻 ack 报文对来估计往返时延的方法，并采用统计的处理方式，首次根据 Netflow 数据中的特定 ack 流记录的起始时间，结束时间，流内报文数等信息进行了往返时延的估计。通过实验表明，该方法具有很好的效果。在 Netflow 技术已经被广泛应用的现代网络，该方法的提出，对于网络管理有着具有重要的意义，为测量延迟提供了一种新思路，新方法。

本文根据TCP流正常传输时的Netflow流记录进行了往返时延的估计，对于数据传输中发生重传或超时情况，则没有给出解决方案。此外，对于利用其它的TCP长流或短流进行往返时延的估计是否具有可行性以及如何进行，本文并没有涉及，下一步的工作将围绕这一主题展开。

6 参考文献

[1] Breibart Y. Chan C Y ，Carofalakis M ，et a1. Efficiently Monitoring Bandwidth and Latency in IP Networks [DB／OL]. www.ieee-infocom.org，2001-04.

[2]Jiang, J, Dovrolis, C. Passive estimation of TCP round-trip times. ACM Computer Communication Review 32 (2002)

[3] 蔡志平，殷建平 ，刘湘辉，吕绍和.一种结合主动和被动方式的网络延迟测量方法[J]. 武汉大学学报（理学版）,2004,50(5): 605-608.

[4] 贾冠昕.利用NETFLOW 技术实现网络流量监测[J]. 信息技术与信息化，2006，5:67-69.

[5] Lindh T. A New Approach to Performance Monitoring in IP Networks Combining Active and Passive Methods [DB/OL] . www.icir.org/vern/imw-2002, November,2002.

[6] 刘 健. Netflow技术原理与应用[J]. 铁道通信信号，2005,41(1):33-35.
[7] 曾嘉,金跃辉,叶小卫. 基于NetFlow的网络异常流量检测[J].微计算机应用,2007,28(7): 709-713.
[8] Netflow Overview. Cisco技术手册.
[9]梁之舜，邓集贤，杨维权等.概率论及数理统计（下）[M].北京：高等教育出版社，1988.