文章编号：无需填写

端到端推测网络路径上的数据包转发优先级

吕国晗, 李星*, 陈焰**

*(清华大学电子工程系新一代网络实验室, 北京 100084)

**(西北大学电子工程和计算机科学系，依利诺依州, 60208, 美国)

摘要:

出于各种的考虑，ISP会对不同类型的数据包设置不同的转发优先级。这会造成不同网络应用出现网络性能上的差异，影响到网络测量结果的代表性，也会增加网络故障排查的难度。通常情况下，ISP不公布此类信息。本文提出了一种端到端测量方法，根据路径上不同类型包丢包率的差异来推断它们是否属于不同的转发优先级。通过在Planet Lab上的测量试验，我们发现了位于某主干网内部的一处优先级设置，结果得到了相关网管的确认。

关键词: 端到端测量; 数据包转发; 优先级

中图分类号: **文献标识码**: A

0 引言

Internet对数据包实现尽力而为的转发。一般情况下，网络并不会对数据包设置不同的转发优先级。但随着不同网络应用的增多，ISP有时会有选择性的改变某类数据包的优先级。据我们所知，在带宽资源有限的情况下，运营商会降低某类数据包的优先级（比如BitTorrent）来限制它们所消耗的带宽；或者提高某类数据包的优先级来保证基于它们的网络应用的性能；或者两者兼有。

一般情况下，ISP并不会公布此类配置信息。但因这种做法违背我们对网络数据包转发的一般假定，如果不了解此类设置的存在，由此类设置引发一些网络现象就变得难以解释，同时也会为我们准确测量网络性能，排查网络故障带来一定的困难。

本文提出了一种使用端到端测量方法，根据路径上不同类型包丢包率的差异来推测路径上是否设置的不同的包转发优先级。本文列举了一处我们在实测中发现的优先级设置。

本文安排如下：第1节是测量和统计分析方法，第2节是测量结果，第3节是结果的验证，最后结论。

1 测量和统计分析方法

1.1 测量方法

发送端使用raw套接字发送ICMP, UDP和TCP类型的数据包，同时将所发包dump下来。接收端在收到一个包后也将它dump下来。在测量结束以后，我们收集到两端dump的文件，由此计算这条路径的单向丢包率。

1.1.1 探测包类型的选择

路由器可以基于数据包头中任一一字段设置优先级，仅TCP源端口号就有65536个。在本文中，我们根据ISP设置优先级可能的目的，有针对性的选择了30种类型（见表1），包括1种ICMP，22种TCP和7种UDP数据包。选择它们是为了探测如下可能：

1．ICMP, UDP和TCP之间是否存在优先级差别；
2．传统的常用TCP应用是否具有高优先级（其中没有80端口的原因是由于我们的测量是在PlanetLab上进行，80端口已被使用）；

收稿日期: 2004-12-07; **修回日期**: 2005-0*-0*.

基金项目:

作者简介: 吕国晗(1978-), 男, 博士生, E-mail: lguohan@gmail.com; 李星(19-), 男, 博士, 教授, 博士生导师, E-mail: xing@cernet.edu.cn ; 陈焰，男，副教授，ychen@cs.northwestern.edu

3．P2P流量是否属于低的优先级。表中所选择的4个端口被eDonkey, Gnutella和BitTorrent所使用；

4．存在安全隐患的端口是否属于低优先级。

表1 30种探测包类型

Tab. 1 30 Tested packet types

PROT	Type/Port Number
ICMP	ICMP_ECHO
TCP	20, 21, 23, 110, 179, 443, 8080 (传统常用应用) 4662, 6346, 6347, 6881 (P2P系统) 161, 135, 137, 139, 445 (安全相关) 1000, 12432, 25942, 38523, 43822, 57845 (随机选取)
UDP	161 (SNMP) 1000, 12432, 25942, 38523, 43822, 57845 (随机选取)

对于TCP和UDP，表中所列端口被做为探测包的源端口。探测包的目的端口为40002。因为ISP很少会特别针对这一目的端口设置特殊的优先级，路由器上不会存在一条针对该端口的显式规则，针对源端口的优先级规则会起作用。因此，我们认为，在大多数情况下这样的探测包能够反映出路由器上对源端口的**优先级设置**。

1.1.2 探测包的发送

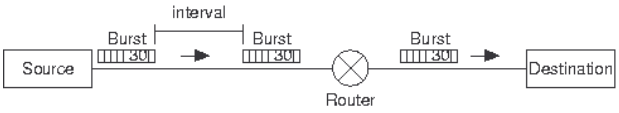

图1 探测包的发送

Fig.1 The probe of tested packets

发送端按轮发送探测包（见图1）。每轮都以随机次序将30个包（每类一个包）背靠背的发送出去。选择按轮发送是为了在发生路由抖动或者设备哆嗦时，一轮的数据包要么全通过要么全不通过，从而避免这些因素导致的不同类型包的丢包率差异。相邻轮之间的发送间隔不能太小，太小会造成相邻轮的丢包出现相关，这违背我们分析时所使用的统计方法的假定。每个包的长度都是100字节（包括IP头）以消除包长可能对丢包率的影响。

1.2 统计分析方法

我们首先使用似然比检测判断一条路径上不同类型包的实际丢包率是否有明显不同。一旦显示有明显不同，我们进一步使用丢包率排序法把这些类型分为几组，其中组内的丢包率没有明显差别，而组间有明显差别。分组就对应了路径上的优先级划分。

1.2.1 似然比检测

假设有k种类型的包，T_i类型的包发送n_i个包，其中l_i个丢失。假设丢包独立等概率，T_i丢包率是p_i，则l_i服从二项分布。我们使用似然比方法作如下假设检验：

H_0：$p_1 = p_2 = ... = p_k$, 所有p_i相等，

H_1：不是所有p_i相等。

由于篇幅的限制，似然比检验的具体方法参见[1]。在实验中，我们取显著性水平α=0.01。拒绝H_0假设的路径被我们称为多优先级路径(MPP)。

1.2.2 丢包率排序法

划分优先级需要使用多个路径测量结果，从它们中找出共同的丢包率排序。首先，我们对每个路径测量结果做丢包率排序。根据排序结果，得到每种包类型的位置号，有i个类型的丢包率严格小于它，它的位置号就是i。它的相对位置是用位置号除以该路径总类型数k减1。在每个路径测量结果中，丢包率最小的类型的相对位置是0，丢包率最大的类型的相对位置是1或者接近1。丢包率相同的类型的相对位置也相同。

然后我们在这多个测量结果中对每种类型的相对位置取平均，得到该类型的平均相对位置。在所测量的路径上，如果一个类型比其它类型的优先级别低，考虑到丢包率有一定随机性，该类型的丢包率在大多数次测量结果中的都应该是最大的，因此平均也应该最大；如果几种类型属于同一优先级别，则在每次测量结果中它们的相对位置应该相同或者随机排列，因此，它们的平均相对位置应该大致相同。

根据平均相对位置排序后画出的曲线，我们可以判断所有类型是否分属不同优先级。如果都是同一优先级，则曲线几乎水平，略微有些上升；如果分优先级，则曲线会有阶跃出现。一般情况下，优先级个数不会超过3~4个，阶跃会很明显。因此，通过观察曲线的阶跃情况，我们可以推测路径上优先级的设置。

2 PlanetLab上的测量和结果

2.1 PlanetLab上测量实验

PlanetLab是由遍布全球的500多台Linux主机组成[2]。我们选取了其中126台主机测量它们两两之间的网络路径。实验从2005-07-01 00:55到2005-07-01 18:16 UTC。每台主机都向其它每台主机发送360轮探测包，相邻轮的平均间隔为3分钟。实验结束后，我们从这些主机收集dump文件，总共获得14,912条路径的测量结果。主机间两个方向的路径认为不同。在测量中，为了让TCP包穿透防火墙，每个TCP包都带SYN标志位。

这里我们只给出与某一网络C的相关测量结果，在所选的126台主机中有两个主机A和B位于该网络内部。根据C网络网管的要求，我们隐去C网络的名称和A, B的IP地址。

2.2 测量结果和分析

与A和B相关的有484条路径，平均丢包率为2%。似然比检测出195条MPP，占40%。表1显示这些MPP的具体分布。可以看出，MPP主要分布在A, B的出路径上。

表1 MPP路径分析

Tab. 1 MPP Path Analysis

路径	总数	MPP	MPP(%)	Loss rate	Match
A->Internet	118	97	82.2	2.5	86
B->Internet	118	80	67.8	4	71
Internet->A	123	9	7.3	1.1	-
Internet->B	123	9	7.3	1.1	-
A->B	1	0	0	0	-
B->A	1	0	0	1.3	-

2.2.1 优先级划分

图2显示了使用丢包率排序法得到的97条A出路径上的平均相对位置曲线。该图横坐标是包的类型(协议号和源端口号)，纵坐标是平均相对位置。曲线有3个明显的台阶，分别对应3个优先级，其中常用TCP应用端口（如telnet, pop3, https等，但不包括P2P应用）和ICMP的优先级最高；其次是UDP包；其它TCP端口的优先级最低。图中横轴最左边的两个点对应被封禁的端口。B的出路径的曲线与A相同，而A, B的入路径没有明显的阶梯出现。

图2得到的优先级划分是一个平均结果，因此，我们想进一步了解每条MPP符合这一划分的情况。由于随机影响，不可能每条路径的丢包率排序都与优先级完全对应，会出现高优先级别的最大丢包率大于低优先级最小丢包率的情况。我们的判断条件是，对于一条路径，当最高优先级中的最大丢包率小于等于最低优先级的最小丢包率时，认为该路径符合推测出来的优先级划分。表2显示对于A, B的出路径，符合率均为88%。对于A出路径，在不符合的11条中，有5条是因为ICMP有最大的丢包率（这可能是受路径上其它路由器的影响），剩余的6条则存在高低优先级丢包率区间重叠的情况。以上结果说明在大多数MPP中，图2所示的优先级划分得到了体现。由于A的很多出路径都共享同一优先级划分，被配置的路由器也很可能位于靠近A的位置。

相反，对于A, B的入路径，不仅MPP在总路径中的比例小，而且也没有同一优先级划分，因此产生这些MPP的路由器应该不靠近A。需要指出，上面的判断条件只对最高和最低优先级进行了比较。这是由于，UDP包的平均丢包率和其它两个优先级的平均丢包率差别不大，因此在很多MPP中UDP的丢包率区间和高低优先级的丢包率区间都有重叠。

2.1.2 推测被配置的路由器

上节大多数A, B出路径都显示出相同的优先级划分。一般来说，这些MPP更有可能是由同一个被配置的路由器，而不是分别由几个配置相同的路由器造成的。

结合traceroute数据，我们可以找出这些MPP从源主机出发共享的最大跳数。而被配置路由器应该在这个最大跳数内。我们发现A的MPP对应的最大跳数是10，B对应的最大跳数是5，但是最大一条对应的路由器都是相同的x.x.x.197。这些共享路径上的路由器都位于C网络内，因此被配置的路由器也应该在C网络内。

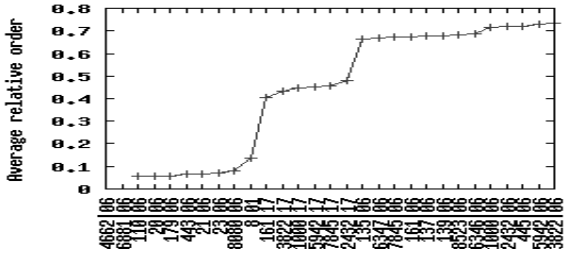

图1 A出路径的相对位置曲线

Fig.1 Relative order curve for outgoing paths from A

3 测量结果的验证

3.1 验证方法

为了验证上面发现的优先级确实是由于路径上某个路由器引起的。我们需要选择几条MPP并逐跳测量从发送端到这些路径上的每台路由器的单向丢包率。当我们发现两种类型的包在某一跳出现明显的丢包率差而这种差异在该跳之后一直存在，则丢包率分叉点就应该是设置了优先级的那台路由器。这种测量很有挑战性，现有的工具如tulip[3]基于TTL超时引发ICMP差错包文，但ICMP发送速率限制使得统计收到的ICMP不能准确的反映由拥塞引起的丢包。

我们的方法如下：首先使用traceroute找出该路径上所有路由器的IP地址。然后向每台路由器发送不同类型的数据包并统计应答包。此时，我们发送TCP SYN包接收RST包；发送ICMP回显请求接收ICMP回显应答；发送UDP包接收ICMP端口不可达差错报文。考虑到后两种应答包也是属于ICMP类型，在本文中，我们只比较TCP之间的不同优先级。

虽然我们的目的是研究，但这种方法针对沿途路由会产生大量的探测包，所以我们只选择4种探测包，每种包向每台路由器发送1000个。

由于该方法实际测量的是发送端到某台路由器的双向丢包率，还需要估计正向路径的丢包率。因此，在选择待验证MPP的目的地址时，我们选择那些A到目的地址的反向路径不是MPP且正反向路由对称的目的地址。这样，反向路径不会引起丢包率差异，差异只由正向路径带来。

3.2 验证结果

我们验证了从B出发的3条MPP。图2是一条MPP的路径丢包率。从5跳开始TCP 1000, 12432和20, 21端口的出现持续的丢包率差，前两个端口对应低优先级，后两者对应高优先级。这与前面推测的优先级划分是吻合的。另外，分叉点的位置显示设置优先级的路由器应该是在第4跳或者第5跳。这与前面推测的最大跳数是吻合的。其余两条路径有相同的结果。

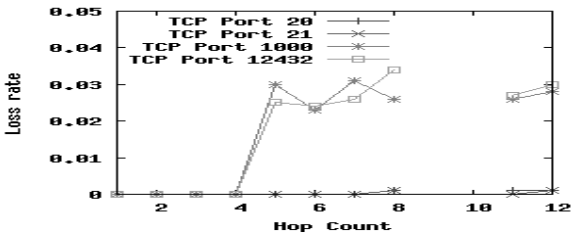

图1 A到143.89.16.12路径的逐条累计丢包率

Fig.1 Cumulative hop-by-hop loss rates from A to 143.89.16.12

4 结论

本文提出了一种通过端到端测量网络路径上的丢包率来推测路径上数据包转发优先级的方法。本文通过在PlanetLab实验床上的测量，发现了在某网络内部的一处优先级设置，并发现很多从该网络到Internet的出路径都受该设置的影响。本文还使用逐跳测量验证了上面的结果，并进一步得到了相关网管的确认。

需要指出即使链路设置了不同优先级，但如果在测量时链路空闲，该方法测量不到丢包率差异，无法探测到优先级的设置。

参考文献

[1] 陆璇, 应用统计. 北京: 清华大学出版社，199
[2] PlanetLab, http://planet-lab.org/php/overview.php
[3] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson. User-level internet path diagnosis. In *ACM SOSP*, 2003.

Inferring Packet Forwarding Priority through End-to-End Measurement

Guohan Lu, Xing Li*, Yan Chen**
*(Dept. of Electronic Engineering, Tsinghua University, Beijing 100084, China)
**(Department of Electrical Engineering and Computer Science, IL 60208, USA)

Abstract: For various purposes, ISPs set different packet forwarding priorities for different packet types, which has wide influence on the application performance, the results of measurement tools and the networking troubleshooting process. However, such information is usually not available to end-users. This paper proposes an end-to-end measurement method to infer the forwarding priorities of different packet types based on their measured loss rates difference. Using this method, the paper presents a configured router within an ISP backbone discovered in our PlanetLab experiment. The inference result has been confirmed by the related network operator.

Key Words: packet forwarding priority, end-to-end measurement, GLRT